United Nations | Department of Economic and Social Affairs

World Population Prospects and 1950 -2020 estimates for age -specific fertility patterns: past experience and future plans

(P. Gerland and G. Gonnella, Population Division)

United Nations Expert Group Meeting on the evaluation of adolescent fertility data and estimates

Session II: Monday 26 October 2020

United Nations | Department of Economic and Social Affairs

Past experience with WPP

5-year age group fertility data and time trends

UN World Population Prospects (WPP)

- **Scope**: 235 countries/areas + > 30 geographical and socio-economic/political aggregates
- **Time**: 5-year periods from 1950 to 2020 (and projection until 2100)
- **Dimensions**: population by 5-year age groups and by sex, fertility by age, mortality and migration by age and sex
  -> 67 indicators (25 by sex, 18 by sex and age) by 5-year periods -> annually *interpolated* subset as by-product
- **Prediction intervals** associated with probabilistic projection, as well as **9 projection scenarios** based 5 fertility variants, 2 mortality variant, 2 migration variants
- **Revision**: every 2-years extensive review / update of past estimates and projections.
- **Online data**: https://population.un.org/wpp/

Aims of the WPP estimates

1. **Comprehensive and standardized demographic dataset for all countries/areas** with internally set of estimates and projections of population size and the three components of population change: fertility, mortality and net international migration

2. **Serve as basis for various projection scenarios** at the global, regional and national level – including derived projections by other international organizations (labor, education, social security benefits, agriculture, health, urbanization, energy, transport, infrastructure, environment, climate change, etc.)

Key concepts for WPP

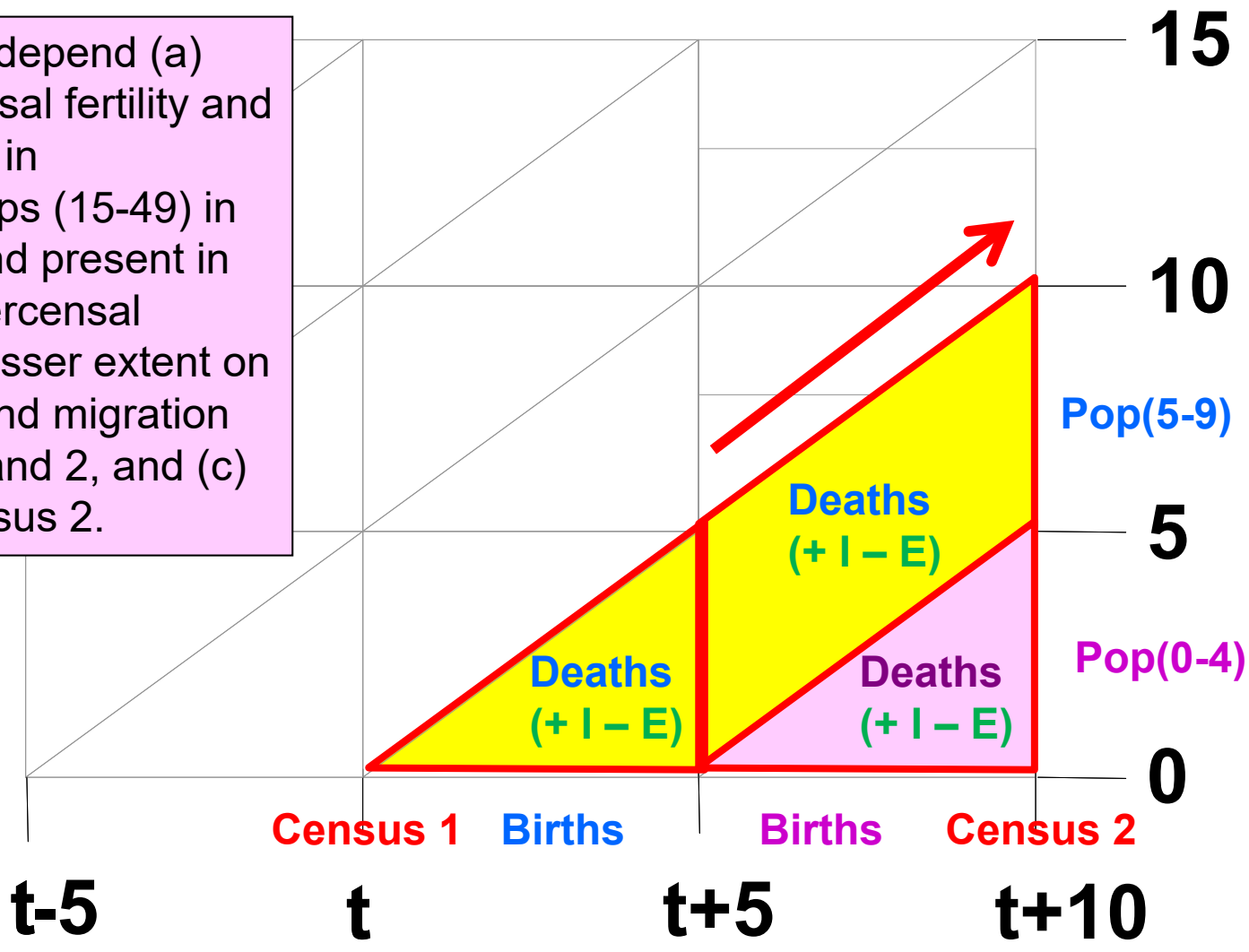
- De-facto vs. de-jure (usual resident) population
- Vital events/rates by year of occurrence
- Population balance (demographic accounting)
- Cohort component
- 5x5 framework -> upgrade to 1x1 for 2021 revision
- Empirical data sources & estimation methods
- Estimate vs. projection
- See *World Population Prospects 2019: Methodology of the United Nations Population Estimates and Projections* for further details

Data requirements for WPP

- For each of the 235 countries or areas:
  - Base population by sex and 5-year age group in 1950
  - For 5-year periods from 1950-2020, time series of:
    - TFR and age-specific fertility rates for women aged 15-49 years by 5-year age group
    - sex ratio at birth (males/females)
    - sex and age-specific mortality rates (life tables) for ages 0-1, 1-4, 5-9, 10-14, …., 90-95, 95-100, 100+
    - net international migration by sex and 5-year age group

Cohort component

Youngest age groups depend (a) mostly on the intercensal fertility and the number of women in reproductive age groups (15-49) in census 1 (surviving and present in the country during intercensal period), and (b) to a lesser extent on infant/child mortality and migration between censuses 1 and 2, and (c) potential errors in census 2.



Data sources (used for WPP 2019)

- National statistical sources (tabulations and/or microdata) either taken as-is or adjusted after in-depth evaluation:
  - **1,690 censuses** (236 since 2010) and post-enumerations surveys
  - **2,700 surveys** (540 since 2010)
  - **vital registration systems** from **163 countries or areas**
  - official statistics reported to the **Demographic Yearbook** of the United Nations
  - **population registers other administrative sources on international migration statistics, education statistics, immunizations, electoral rolls, etc.**

Data sources (continued)

- Refugee statistics from the Office of the UN High Commissioner for Refugees
- Estimated time series of adult HIV prevalence and coverage of antiretroviral treatment from UNAIDS
- Estimated time series of infant and under-five mortality from the UN Inter-Agency Group for Child Mortality Estimation
- Estimates of international migration flows and stocks of foreign-born persons from the UN
- Various other series of international estimates produced by international and regional organizations and academic research institutions

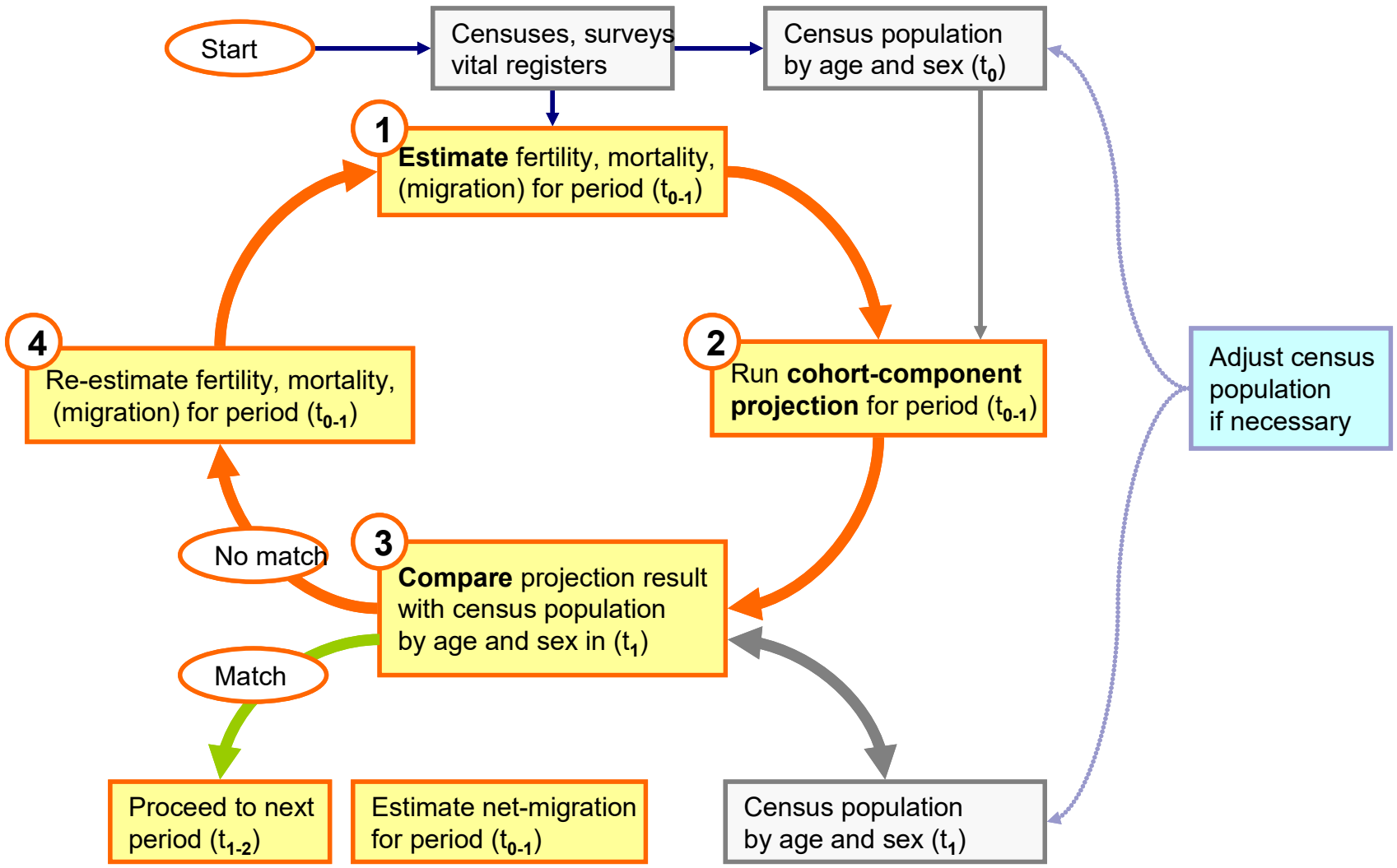
With so many data available, why estimates are necessary…

- **To fill-in gaps in missing data**: most information often available only for some countries and/or dates, or not sufficiently disaggregated by age
- **To reconcile differences** between (a) data sources and/or estimation method(s) for a specific date and (b) within sources over time
- **To ensure international comparability** using similar definitions/concepts, methodology and assumptions across countries

WPP workflow process with TFR & ASFR5

- Compile and compute direct and indirect fertility estimates from as many empirical data sources as possible for each country since 1950
- Review and assess the various series
- Generate an initial robust time trend for TFR (15-49) and ASFR5
- Use this initial set of estimates within the full cohort-component population reconstruction by age and sex since 1950
- Compare and assess the reconstructed population cohorts with those enumerated across the various censuses
- Revise and adjust the set of WPP estimates to reconcile the various demographic components (e.g., TFR) that satisfy the demographic balancing relationships over time, age and cohorts

WPP estimation process for each country/area



Estimation of robust time series for demographic rates

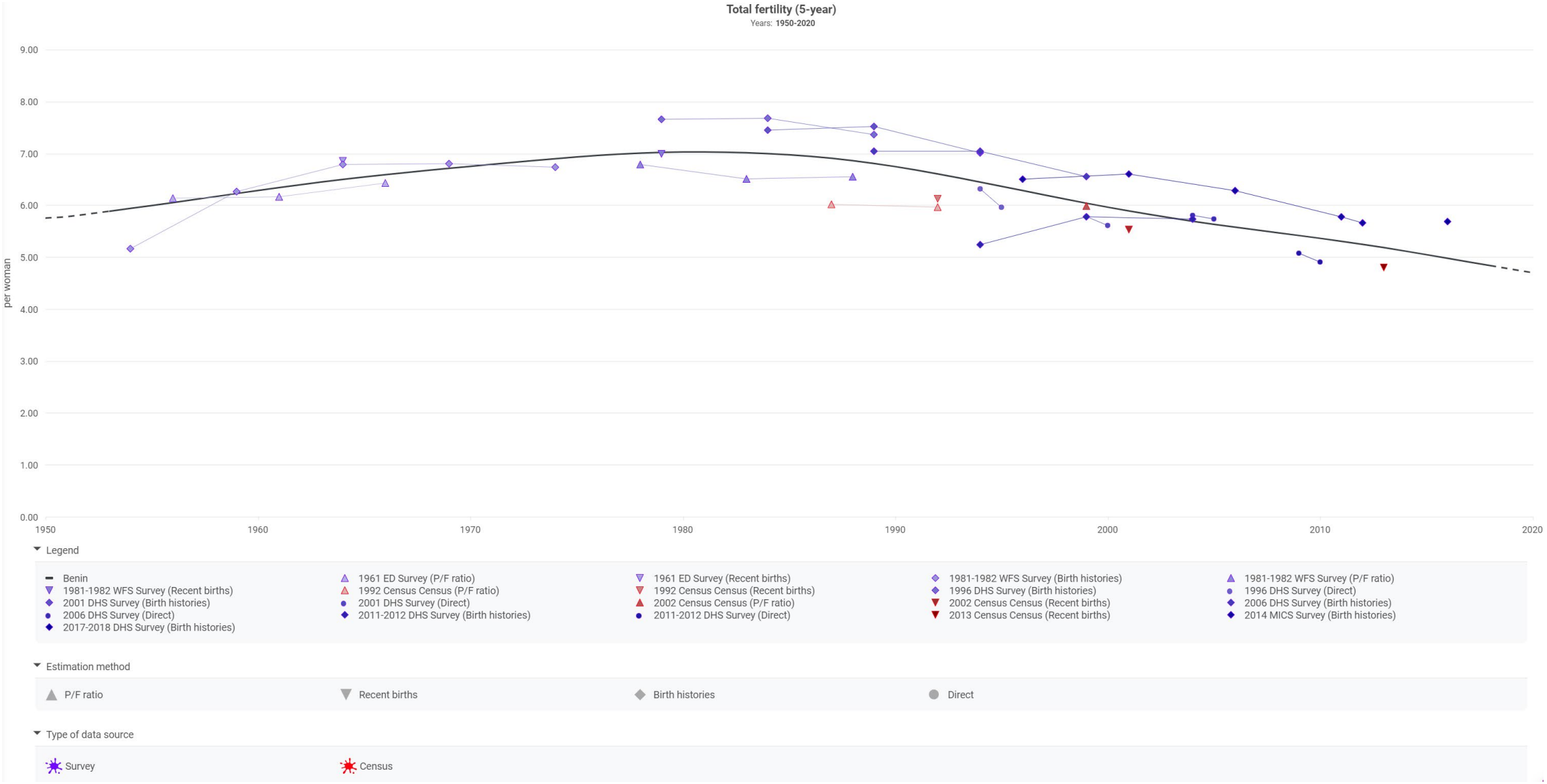
- For many countries, data available vary greatly in quantity, frequency, quality, reliability and consistency.
- Not all data points are as informative and can be trusted equally…
- Estimates can vary based on the **type of data sources** (census, surveys, vital registration), the **type of survey** itself (national survey vs. international survey programs), the **estimation methods** (direct or indirect estimates) and by **various biases affecting reporting of retrospective birth histories or lifetime fertility**.

Sources of data and estimation methods

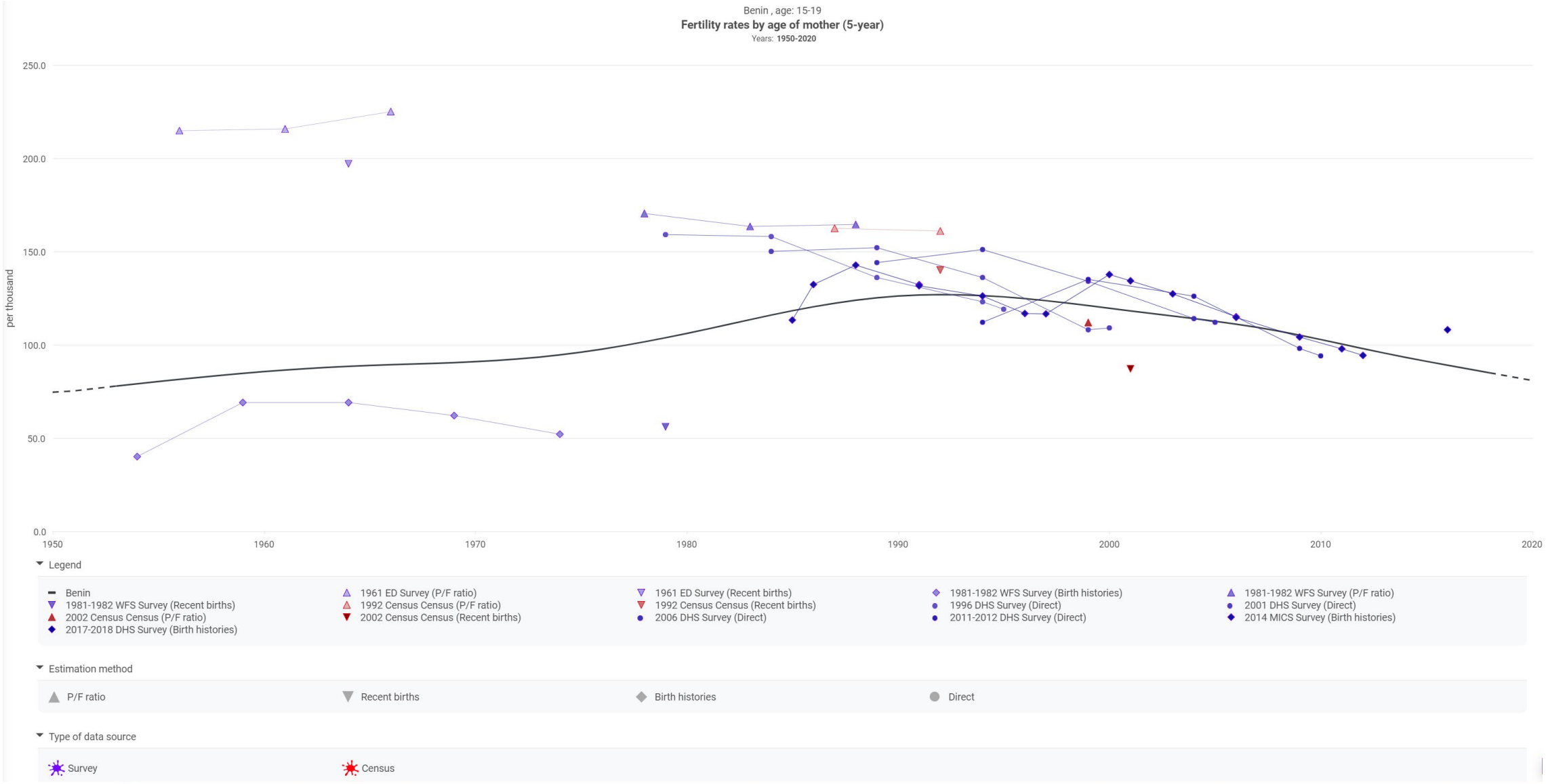
Source	Method	Time period	TFR	ASFR
Official figures	Estimates	**Annual**	✓✓✓✓	✓✓✓✓
Vital statistics from civil registration	Computed rates from DYB-NSO	**Annual**	✓✓✓✓	✓✓✓✓
Surveys	Birth histories (and extrapolations)	**Prior 15-35 years**	✓✓✓✓	✓✓✓✓
Censuses/Surveys	Recent births	**Prior 12-24 months**	✓✓✓	✓✓✓
Censuses/Surveys	Recent births and average parity methods	**Prior 12-24 months**	✓✓✓	✓✓✓
Censuses/Surveys	Children ever born methods	**15-45 years before**	✓✓	
Censuses/Surveys/ admin. stats	Population methods	**Prior 15 years**	✓✓[1,2]	✓✓[1]
Model-based	Other methods[3]	**Prior 5-15 years**	✓	

(1) Own-children method, (2) Reverse survival method, Education stats
(3) Intercensal estimates, regression methods, stable population

TFR: Benin

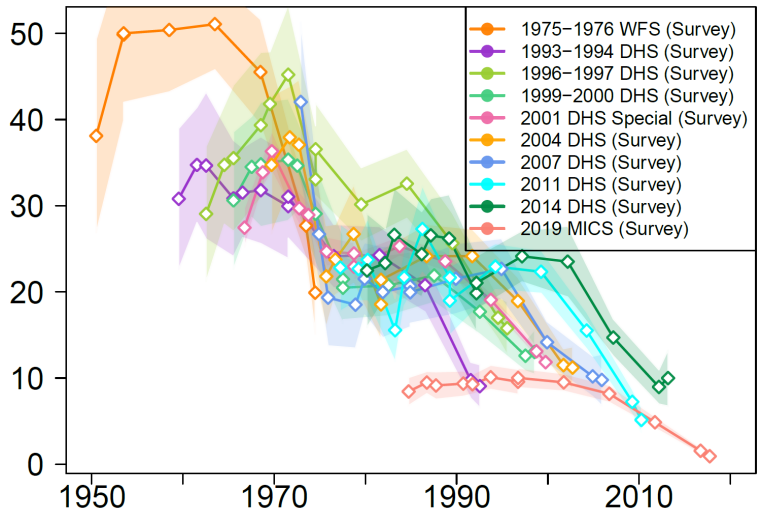

Total fertility (5-year)
Years: 1950-2020
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
per woman
1950
1960
1970
1980
1990
2000
2010
2020
Legend
Benin
1981-1982 WFS Survey (Recent births)
2001 DHS Survey (Birth histories)
2006 DHS Survey (Direct)
2017-2018 DHS Survey (Birth histories)
1961 ED Survey (P/F ratio)
1992 Census Census (P/F ratio)
2001 DHS Survey (Direct)
2011-2012 DHS Survey (Birth histories)
1961 ED Survey (Recent births)
1992 Census Census (Recent births)
2002 Census Census (P/F ratio)
2011-2012 DHS Survey (Direct)
1981-1982 WFS Survey (Birth histories)
1996 DHS Survey (Birth histories)
2002 Census Census (Recent births)
2013 Census Census (Recent births)
1981-1982 WFS Survey (P/F ratio)
1996 DHS Survey (Direct)
2006 DHS Survey (Birth histories)
2014 MICS Survey (Birth histories)
Estimation method
P/F ratio
Recent births
Birth histories
Direct
Type of data source
Survey
Census

ASFR 15-19: Benin

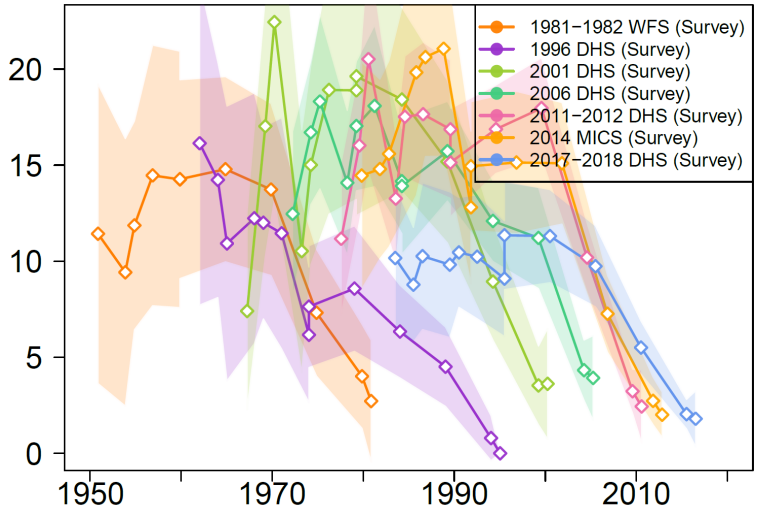


ASFR 10-14 selected empirical series

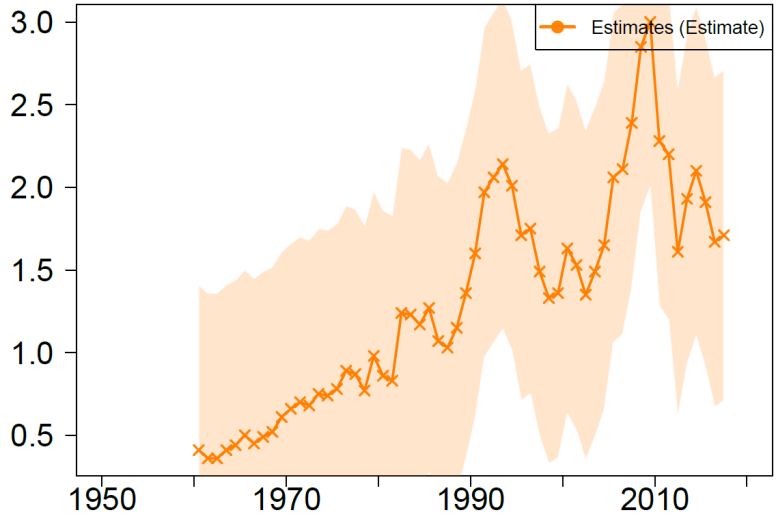
Bangladesh



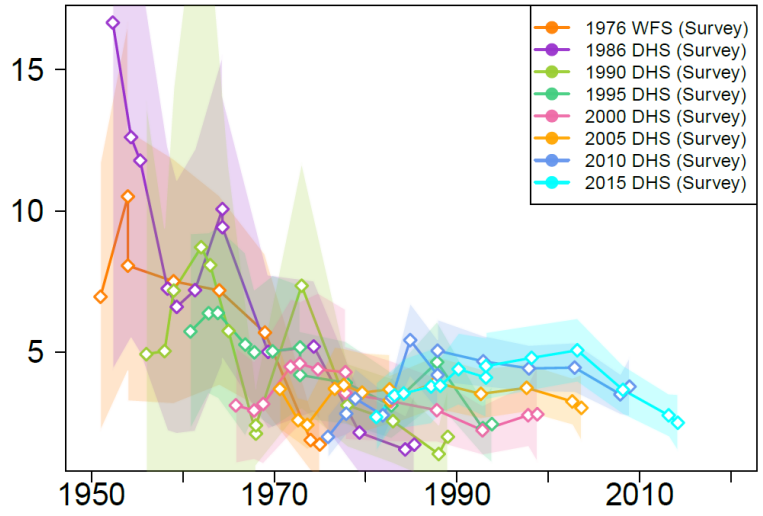
Benin



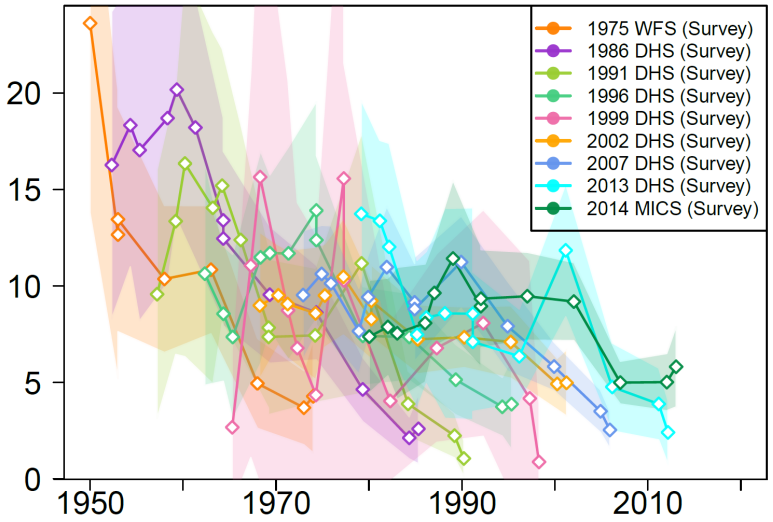
Bulgaria



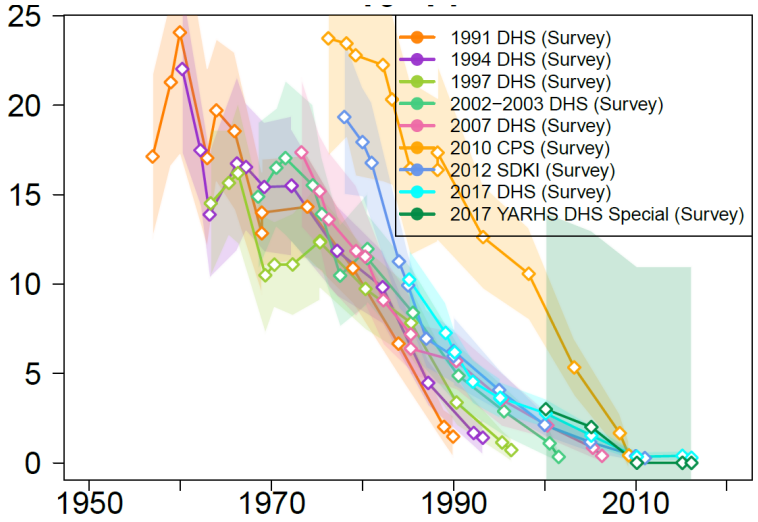
Colombia



Dominican Republic

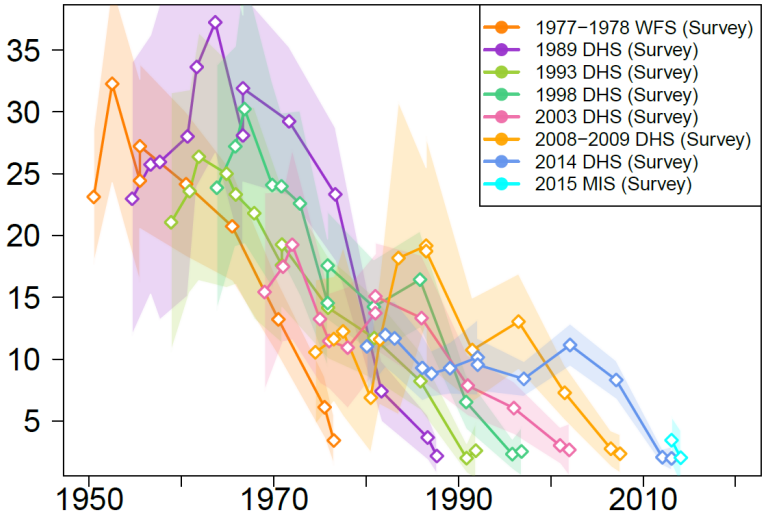


Indonesia

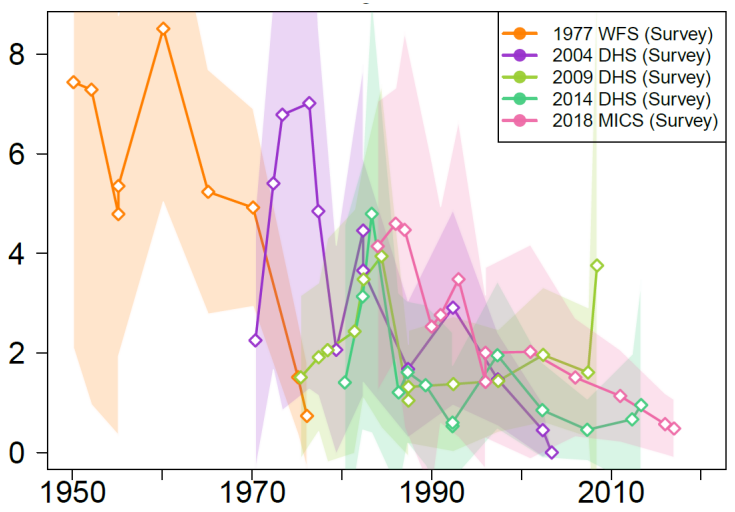


ASFR 10-14 selected empirical series

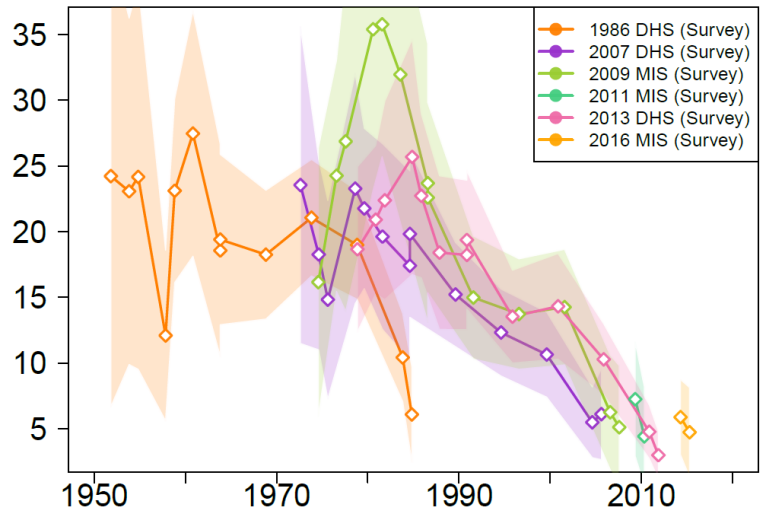
Kenya



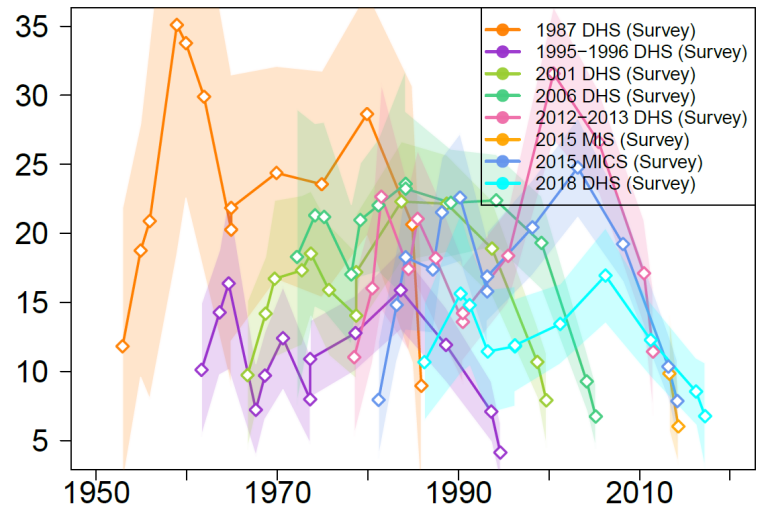
Lesotho



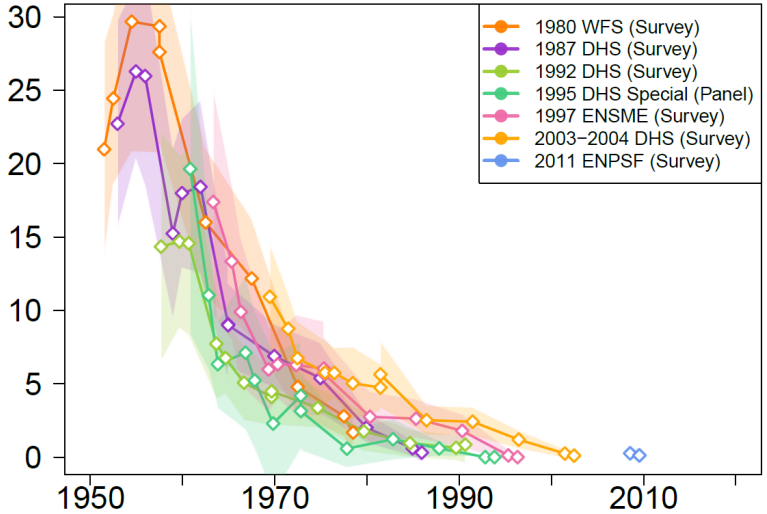
Liberia



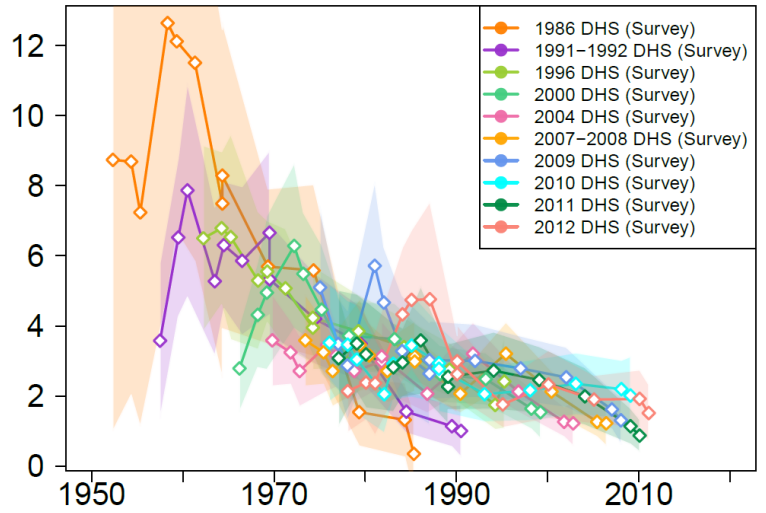
Mali



Morocco



Peru



United Nations | Department of Economic and Social Affairs

Future plans for WPP 2021

(Annual time trends and) single-age fertility patterns

Changes for WPP 2021

- Upgrade production system to **single year and single age data model**
- Improve capacity to **use annual time series** (upon data availability and reliability) including **ASFR for age groups 10-14 and 50-54**.
- Improve capacity to **use single age data** (upon availability and reliability): use for good VR countries, for the rest **use 5-year age groups graduated into single age** using a recalibrated spline model developed by Schmertmann.
- Streamline/harmonize steps used to prepare country data and WPP estimates
- Greater documentation and explanations of the various methods used to derive demographic estimates for each demographic components and the reconciliation with population estimates **->** WPP method protocol
- Provide access to both WPP estimates and underlying empirical data for key demographic indicators -> Data Portal + Demo Data + Data Archive
- More GATHER compliant
- *See Expert group meeting on methods for the World Population Prospects 2021 and beyond* (6-8 April 2020)

Availability of single age fertility data

Over 4,500 series:

- **Vital registration**
  - more than 4,000 annual series for 71 countries covering age 12-55 from 1891 to 2018,
  - Sources are Human Fertility Database (HFD), Eurostat and Human Fertility Collection (HFC) in a hierarchical order (no overlap for each country x year).
- **Survey**
  - 451 series for 109 countries covering age 10- 49 from 1964 to 2019,
  - Sources are DHS, MICS and other surveys collecting Full Birth Histories,
  - Rates for 10 years before each survey, computed by B. Schoumaker directly from micro-data using his Stata code,
- **Health and Demographic Surveillance System**
  - 72 series for 14 countries covering age 10-54 from 1976 to 2018,
  - Rates for 3 to 8 years period, computed by UNPD using the Stata code developed by B. Schoumaker.

Availability by SDG region

Years of observation:

- VR = single year,
- Survey/HDSS = years covered retrospectively

Number of years of observation by time period and SDG region

Sustainable Development Goal (SDG) regions	Before 1950	1950-1969	1970-1989	1990-2009	2010-2019
Sub-Saharan Africa		11	393	1595	680
Northern Africa and Western Asia		26	283	484	102
Central and Southern Asia		9	71	277	121
Eastern and South-Eastern Asia	5	27	125	297	102
Latin America and the Caribbean		42	330	552	78
Australia and New Zealand	42	40	40	40	18
Oceania*		6	5	5	17
Europe and Northern America	336	782	999	1045	437

Number of series by source and SDG region

	Survey			Health and Demographic Surveillance System	Vital Registration***		
Sustainable Development Goal (SDG) regions	DHS	MICS	Other surveys**	HDSS	HFD	Eurostat	HFC
Sub-Saharan Africa	142	28	48	66			
Northern Africa and Western Asia	31	10	20	0	27	67	130
Central and Southern Asia	30	5	6	3			7
Eastern and South-Eastern Asia	25	3	3	3	130		61
Latin America and the Caribbean	49	6	32		14		31
Australia and New Zealand							180
Oceania*	1	1	1				
Europe and Northern America	4	2	4		2402	308	779

Series:

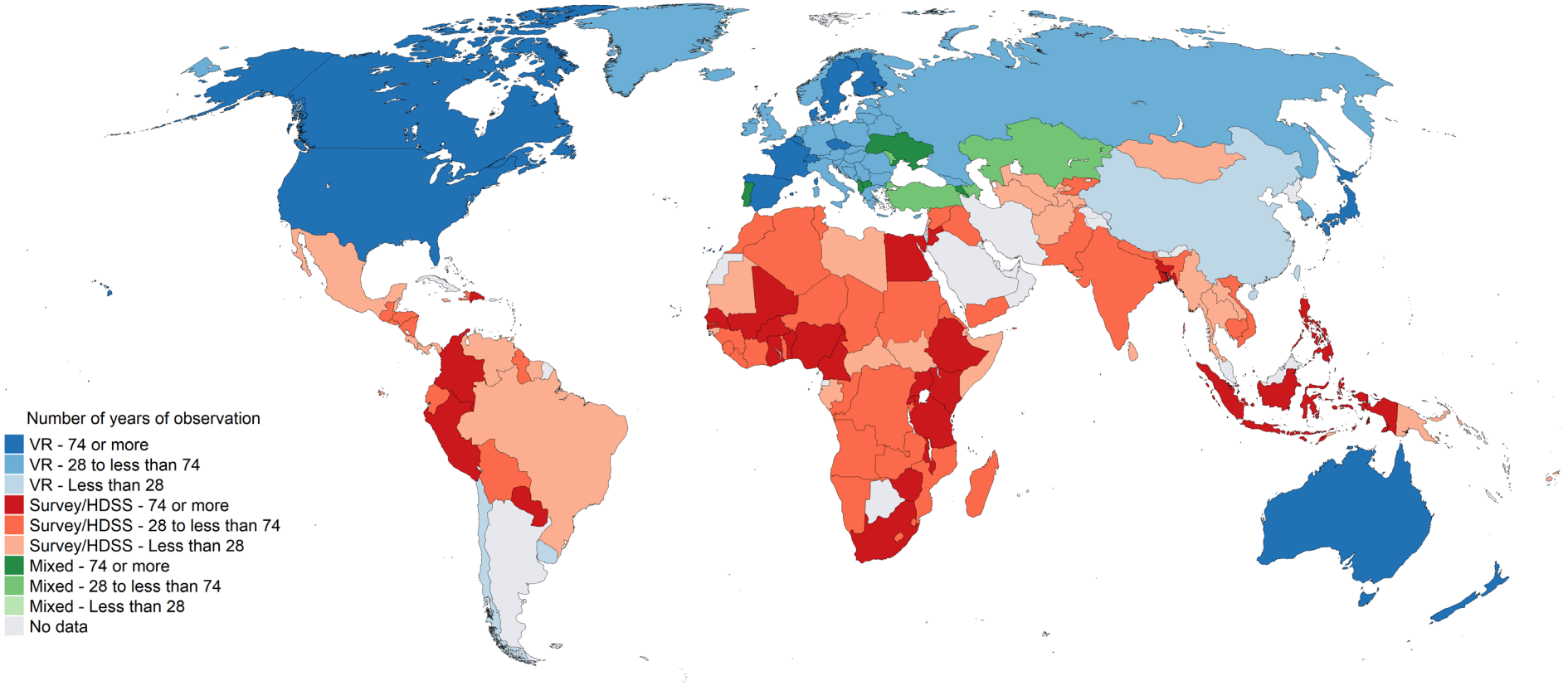
- 4,136 VR,
- 451 surveys,
- 72 HDSS

*(excluding Australia and New Zealand)

**including WFS, MIS, RHS, PHS, PAPFAM, PAPCHILD and more.

***the numbers do not reflect the data availability in the three databases. The criteria chosen is to have one series of Vital Registration data for each country x year. The priority has been given to data coming from HFD, then Eurostat and eventually HFC.

Years of observation by country



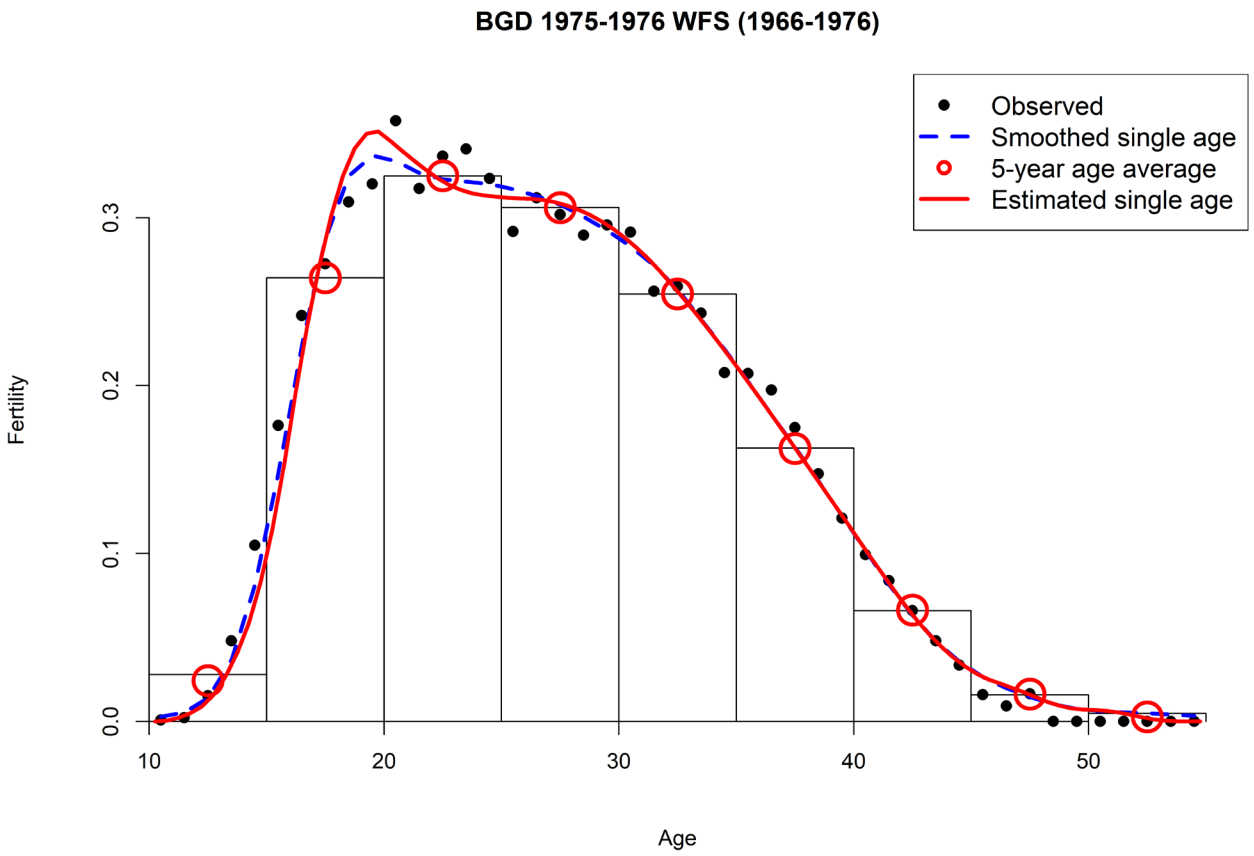
Use of single age fertility data (1)

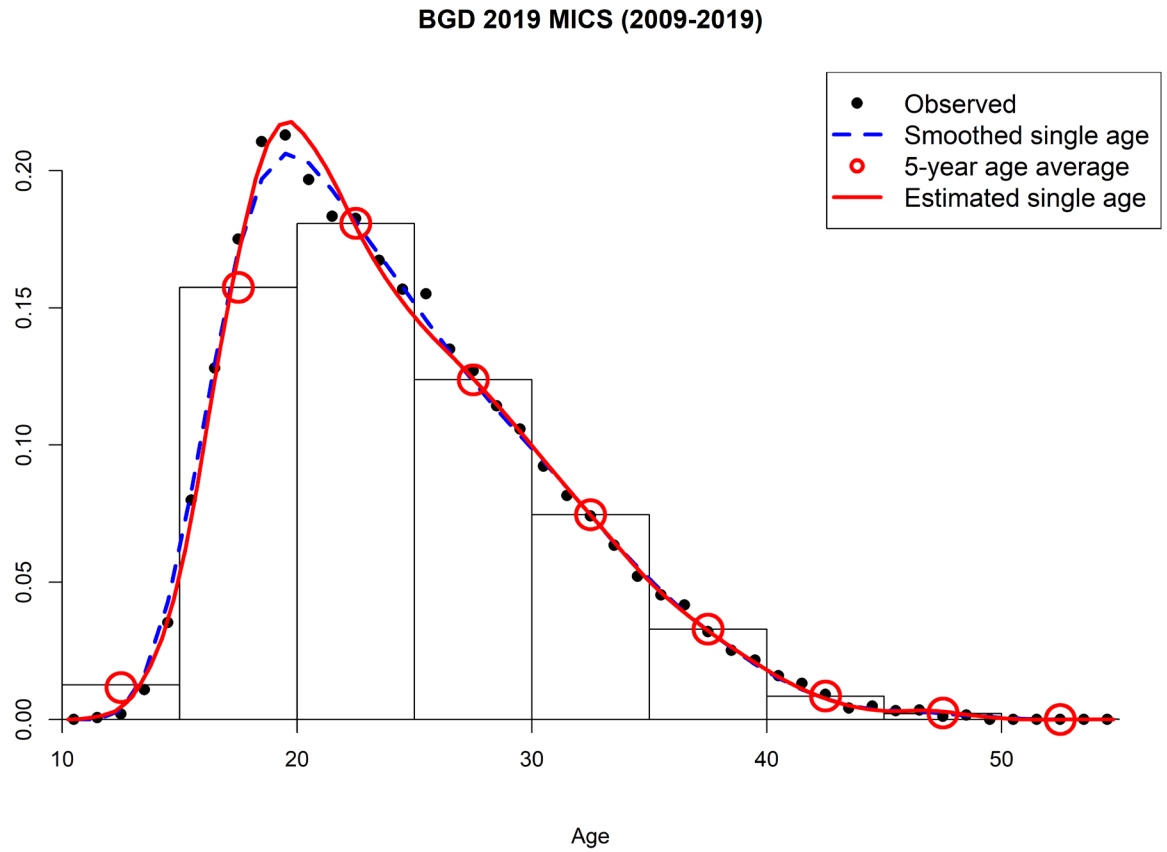
- For countries with reliable **vital registration**, single age fertility series will be used as empirical data;
- For countries that heavily rely on **surveys/HDSS** to gather information on fertility, the series will be **smoothed** using the method as proposed by **Bruno Schoumaker (2020)** and based on **Pantazis and Clark (2018)**:
  - Principal component analysis applied to the 523 series,
  - Single age fertility rates smoothed using a linear combination of the first 5 components resulting from the PCA (capturing 99% of the variance),
  - Data further smoothed using a cubic spline (degree of smoothness determined by cross-validation) with monotonicity constraints on the tails (age < 15 & age 50+),
  - Smoothed series to be used as empirical data.

Use of single age fertility data (2)

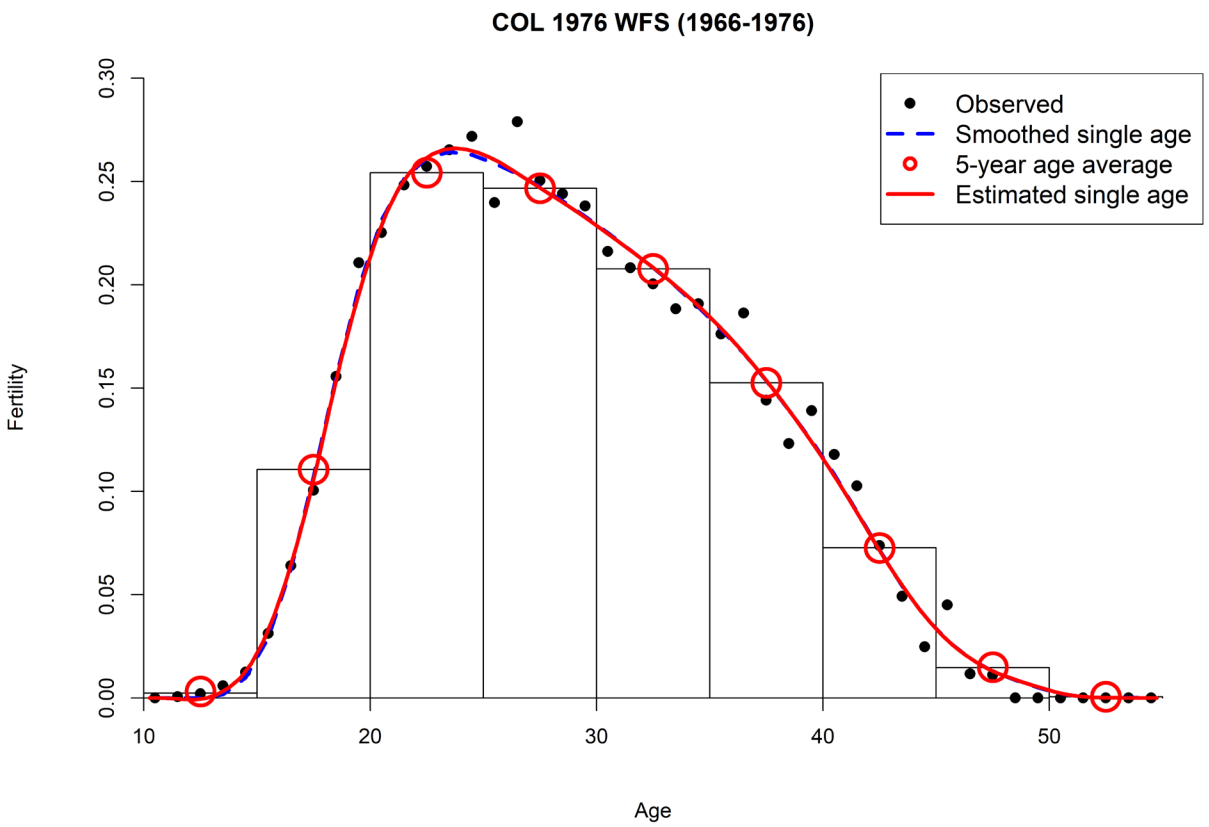
- Information on single age fertility rates to be used for the **graduation** of 5-year age to single age rates using the **Calibrated Spline (CS)** estimator developed by **Carl Schmertmann (2014)**:
  - The method expands observed abridged fertility schedules based on similarity with known single age fertility rates,
  - The output is a set of multipliers that can be applied to any abridged ASFR series to obtain the desired graduated series.
- For WPP 2021, re-calibration of the CS using the available single age series:
  - Vital registration series (5-year average to avoid overcounting of highly correlated data),
  - Survey/HDSS data smoothed using the PCA and cubic spline.

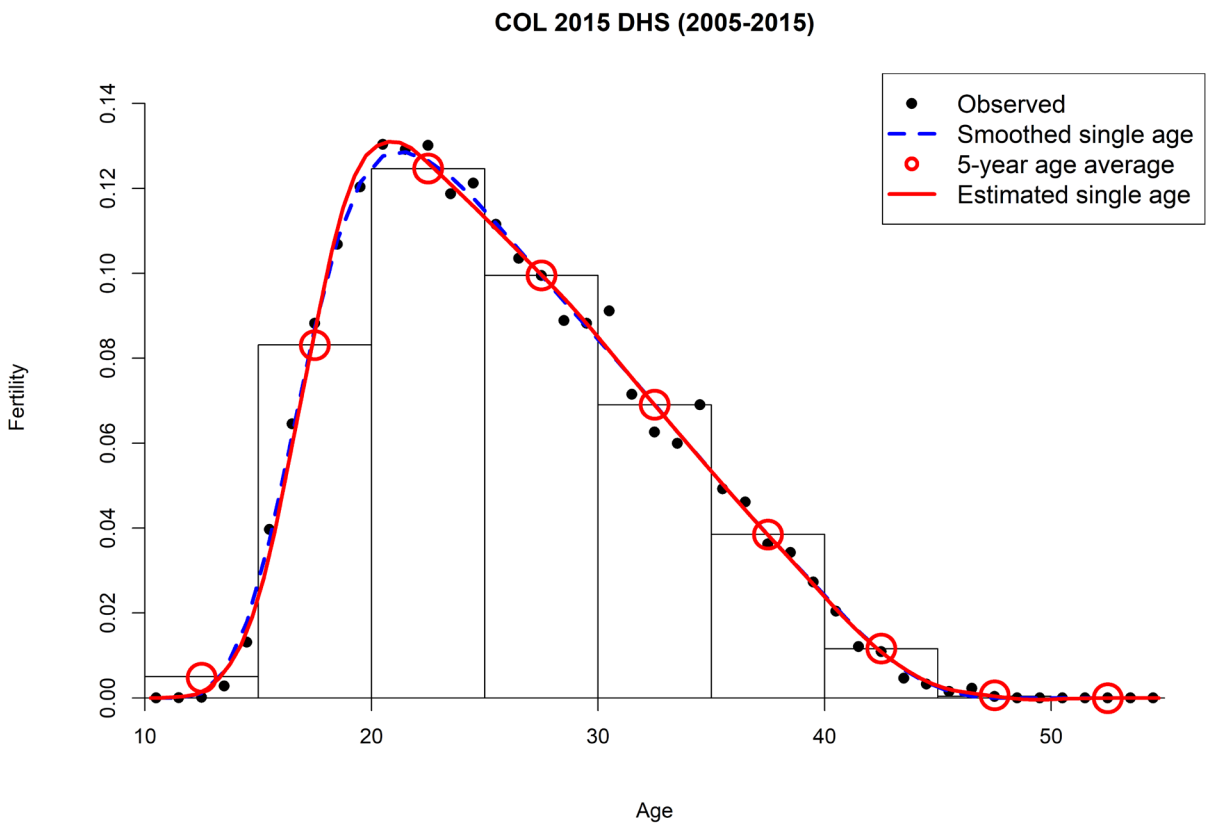
Example: Bangladesh



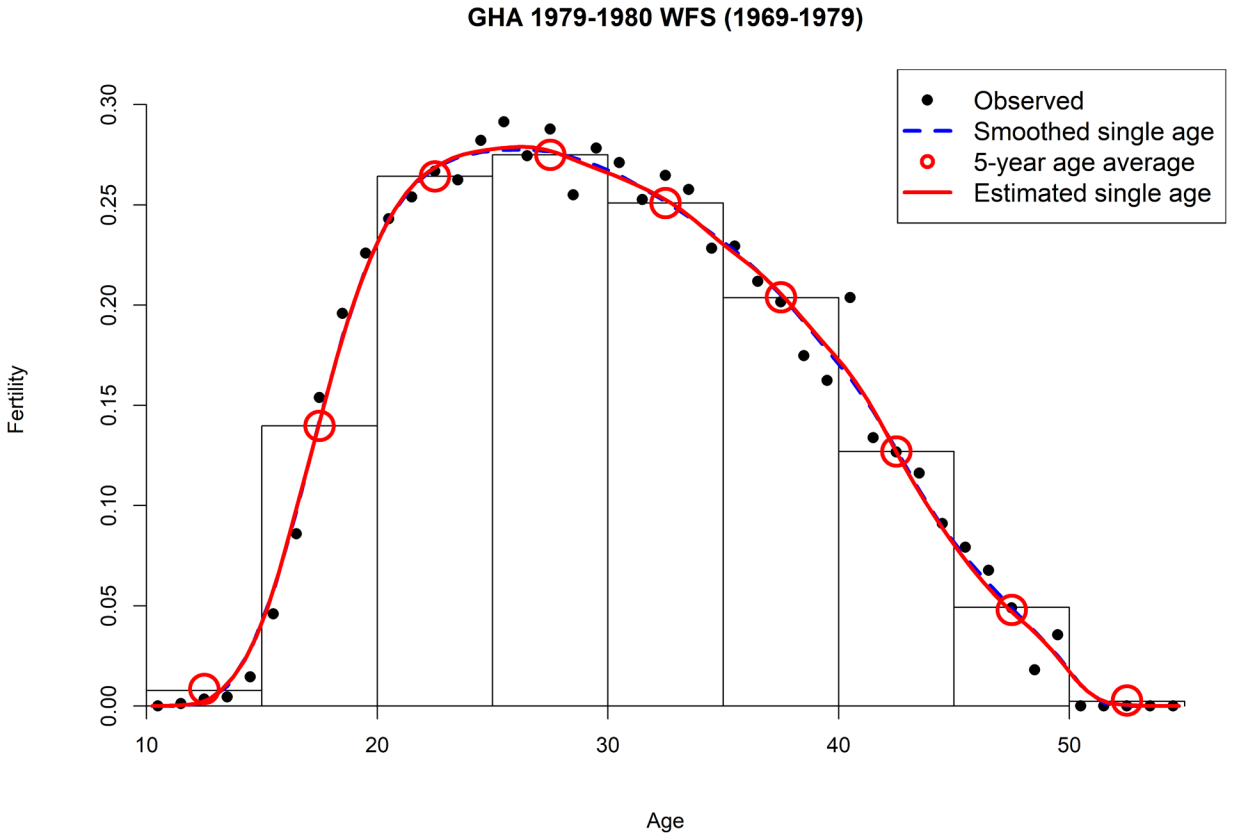


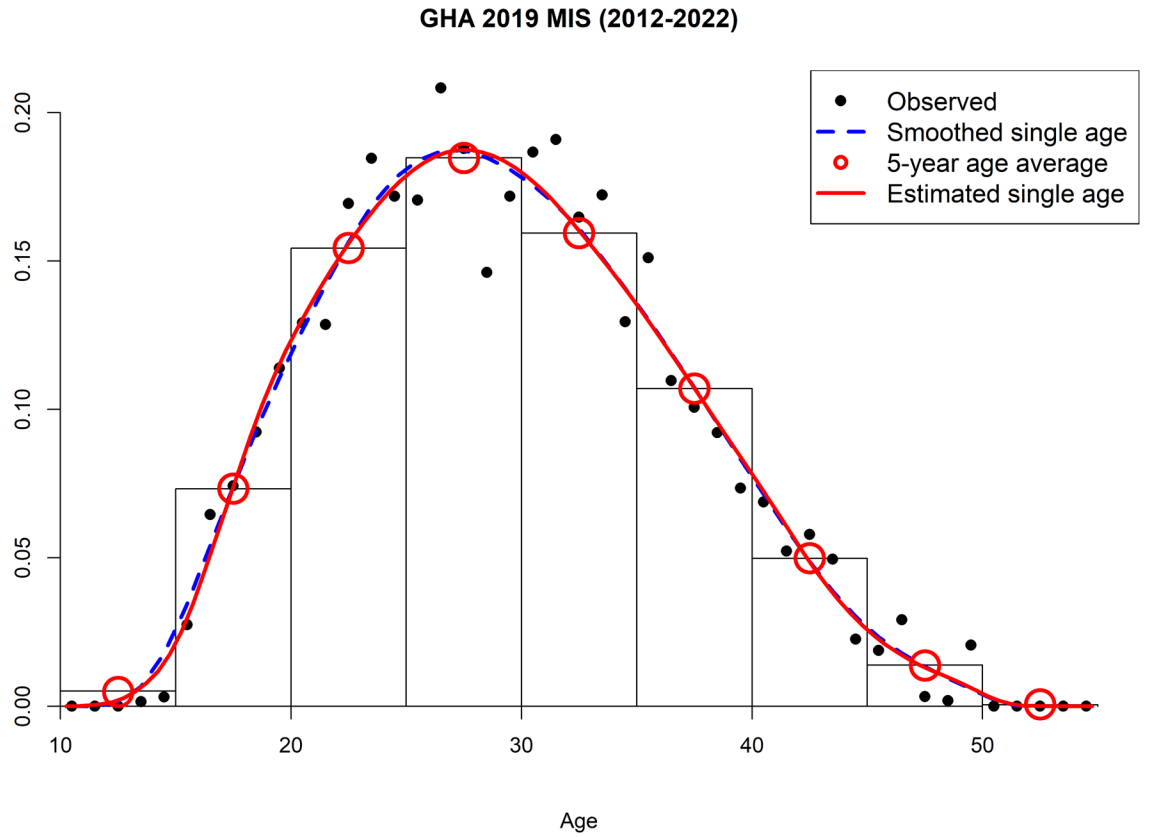
Example: Colombia



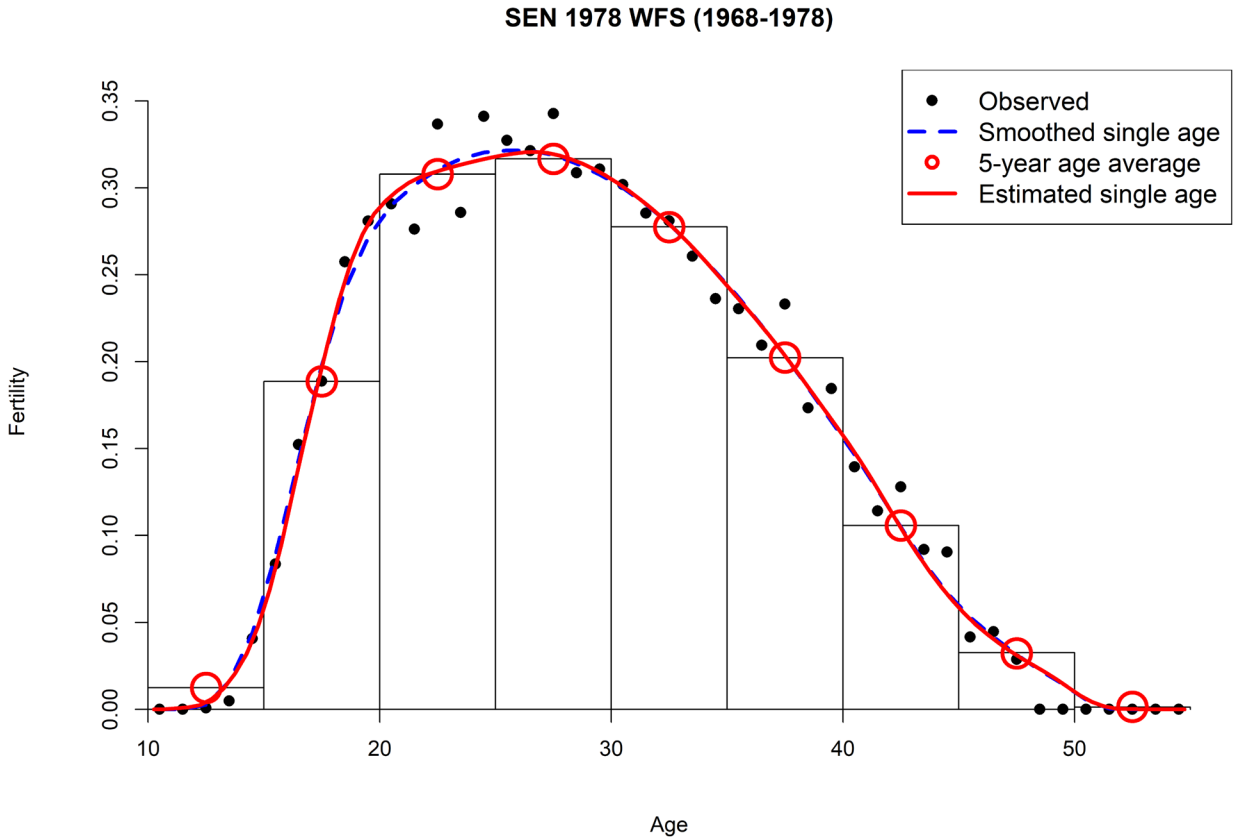


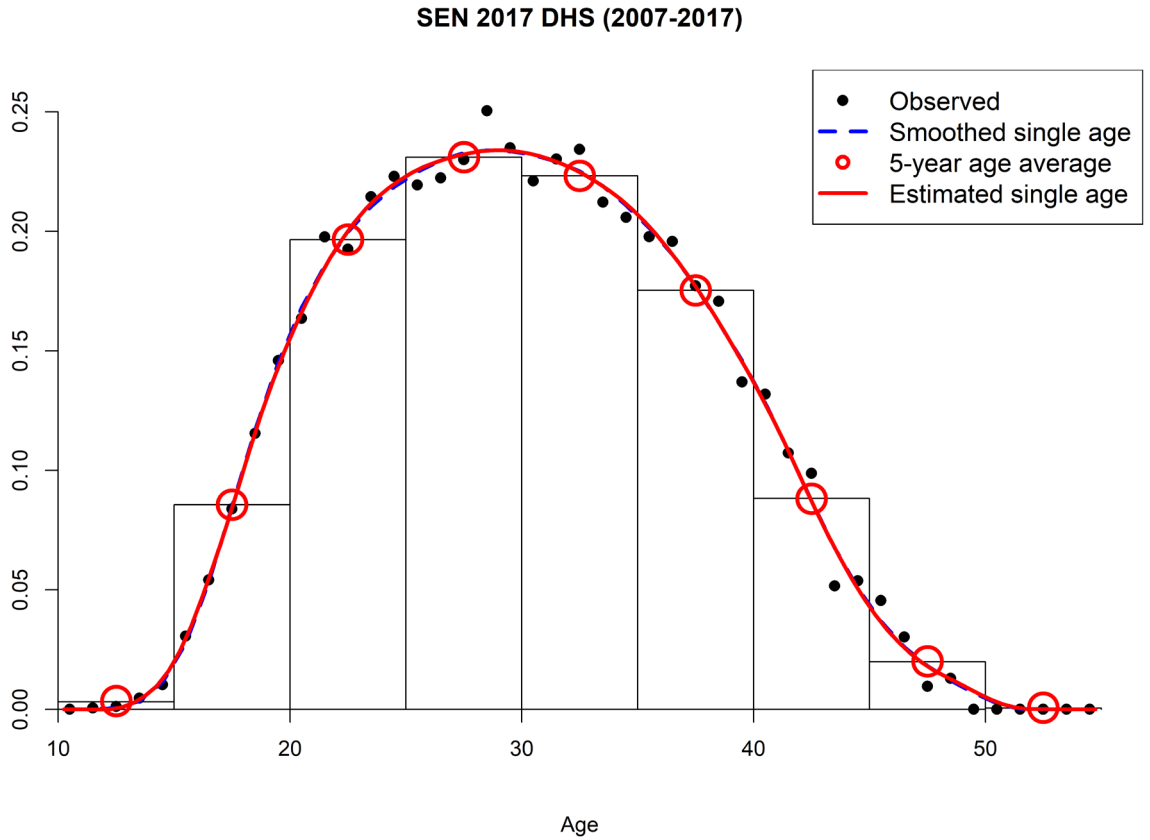
Example: Ghana



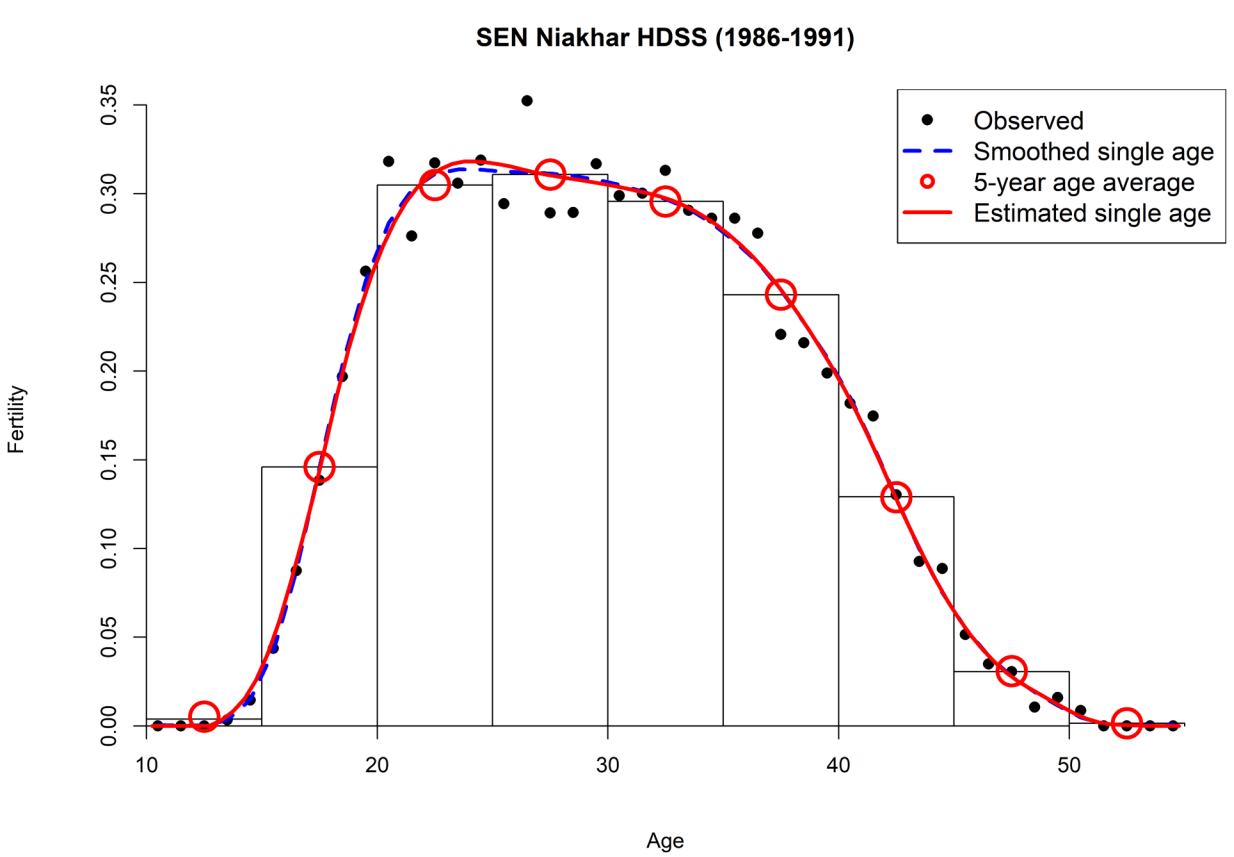


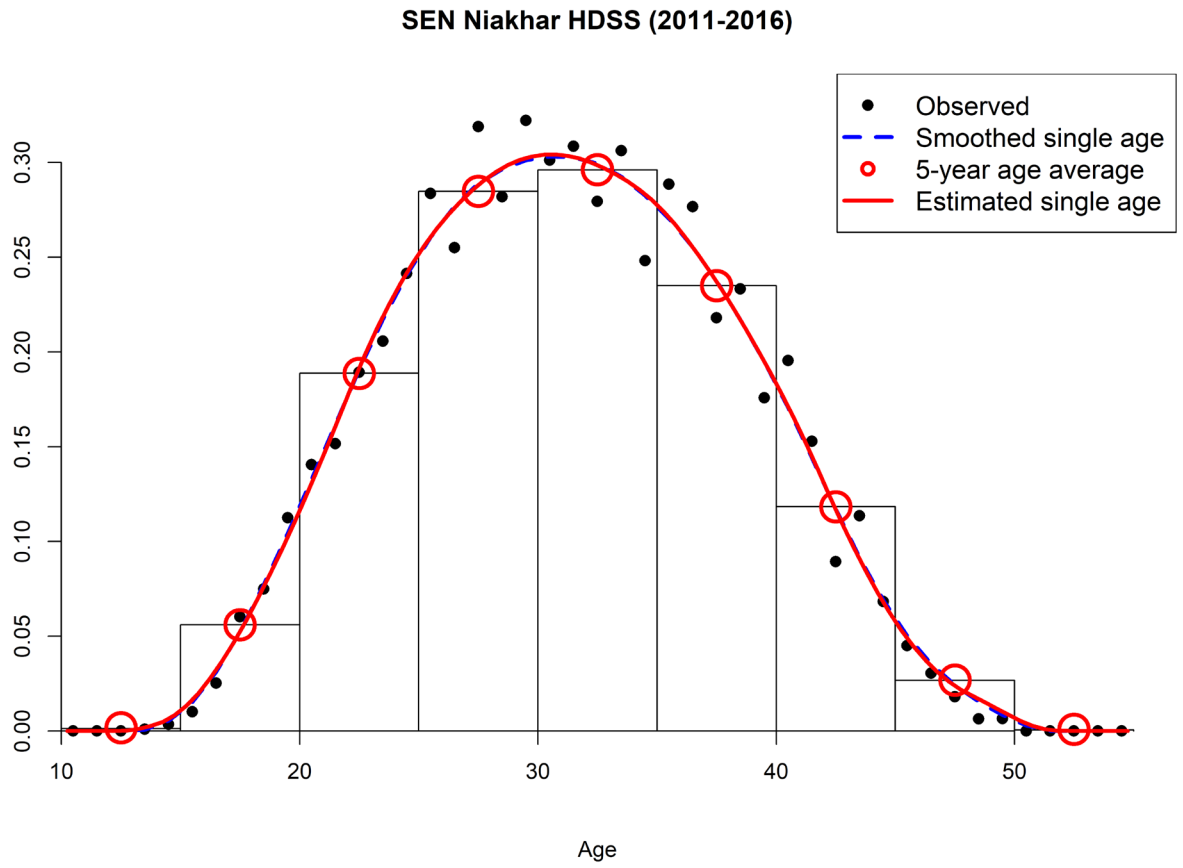
Example: Senegal



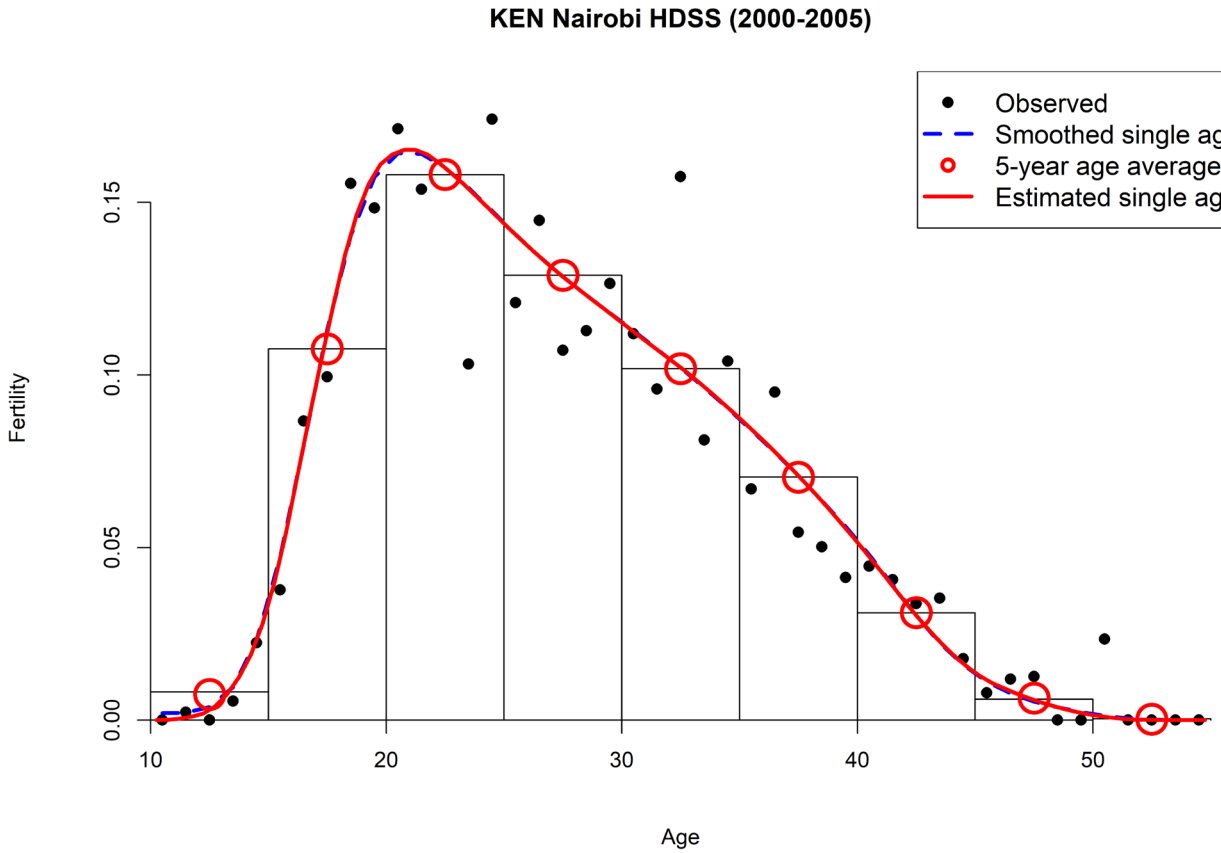


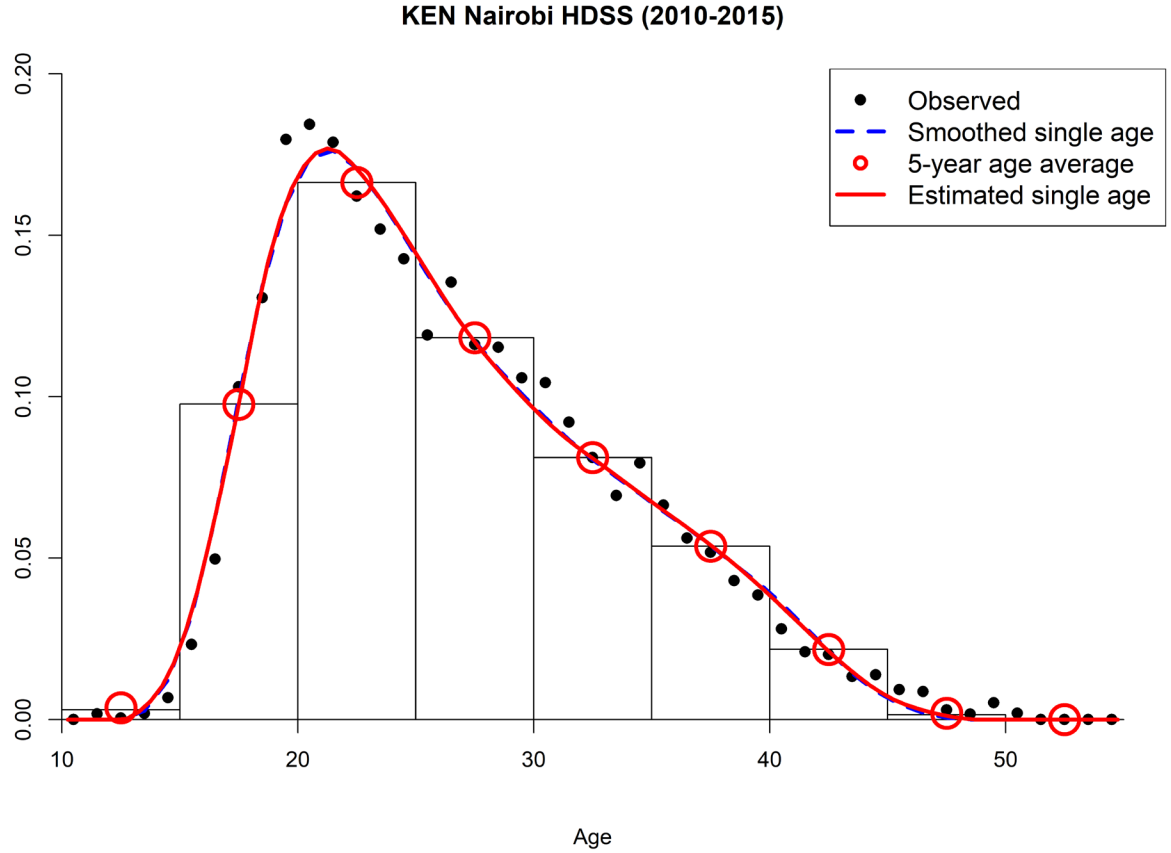
Example: Senegal HDSS



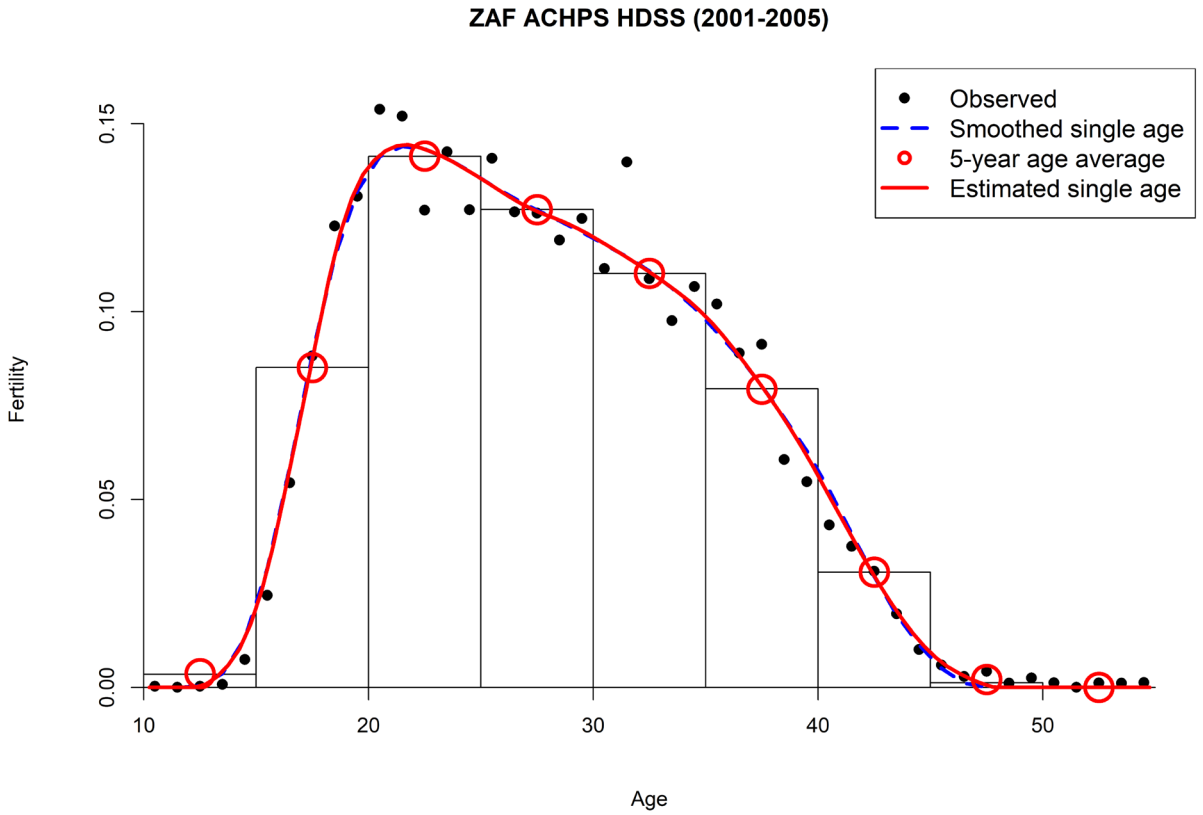


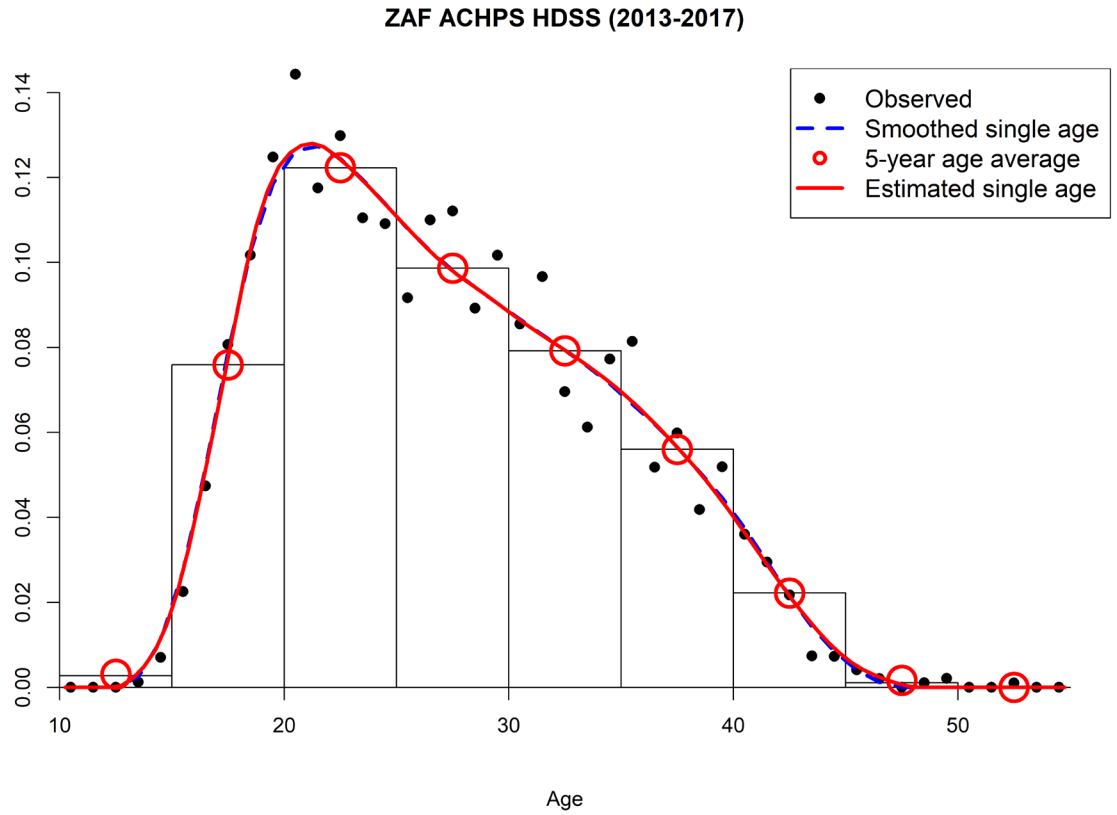
Example: Kenya HDSS





Example: South Africa HDSS





Adolescent fertility

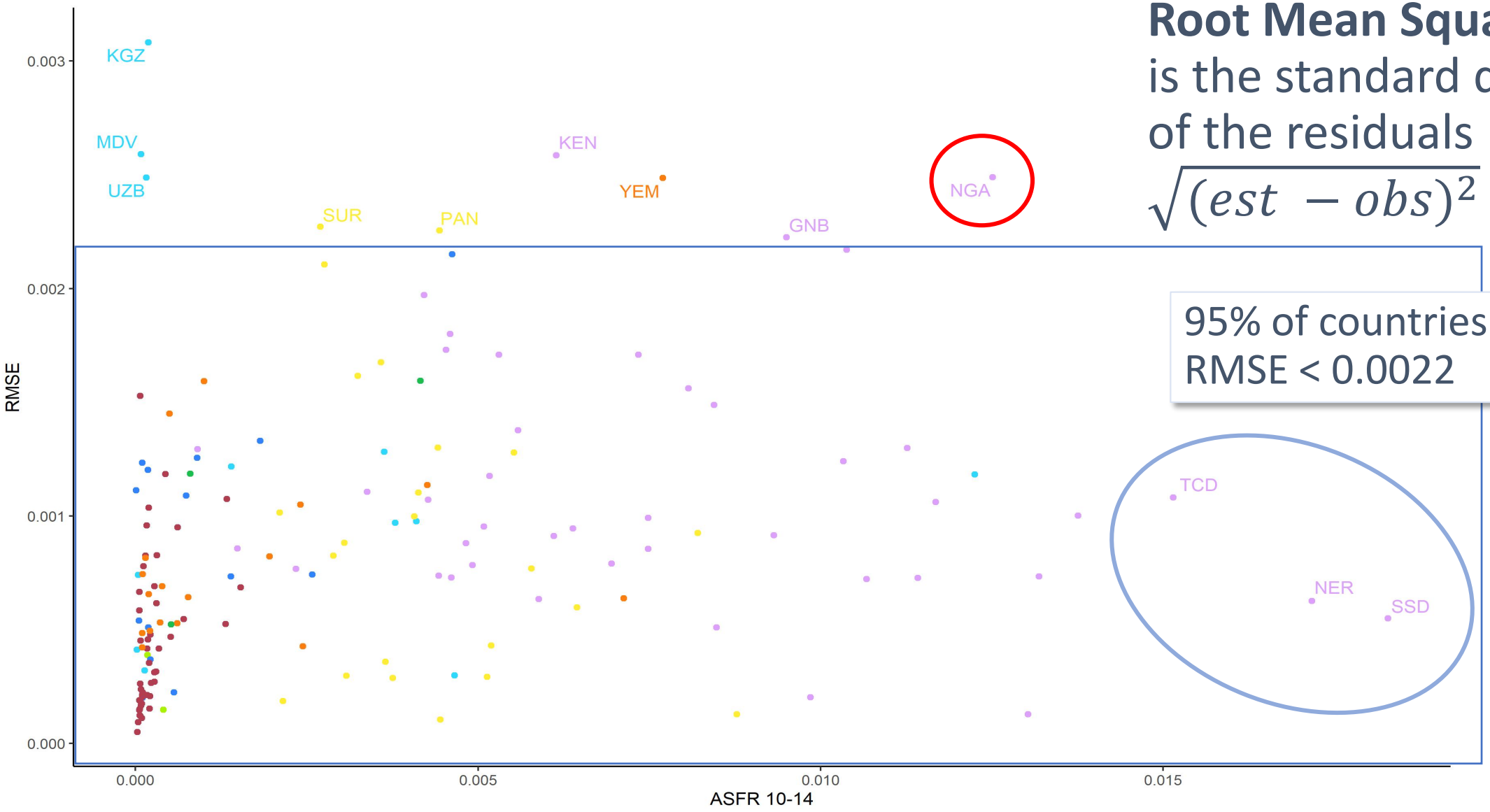
- Vital registration:
  - Age 12-13 available in 2573 series out of 4136, age 14 in 3644, Age 15 in 4127.
- Survey:
  - Age 10-14 available in 446 out of 451 series.
- Health and Demographic Surveillance System:
  - Age 10-14 available in all the 72 series

	Surveys	HDSS	Vital registration		
Sustainable Development Goal (SDG) regions	**Age 10-14**	**Age 10-14**	**Age 12**	**Age 13**	**Age 14**
Sub-Saharan Africa	218	66			
Northern Africa and Western Asia	61		27	27	157
Central and Southern Asia	41	3			
Eastern and South-Eastern Asia	31	3	130	130	182
Latin America and the Caribbean	82		14	14	45
Australia and New Zealand					180
Oceania (excluding Australia and New Zealand)	3				
Europe and Northern America	10		2402	2402	3080

Model performance on age 10-14 by country

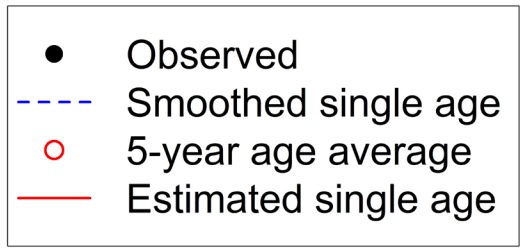
Root Mean Square Error is the standard deviation of the residuals

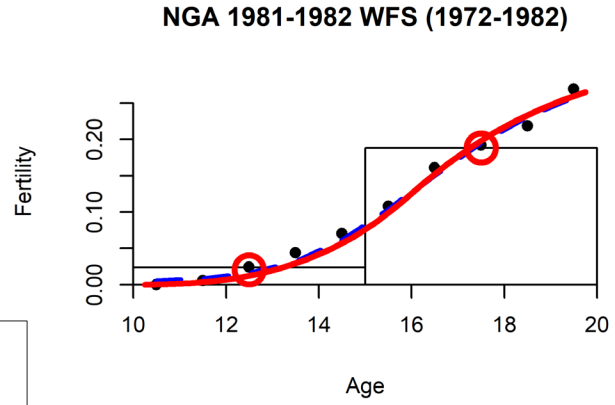
$$\sqrt{(est - obs)^2}$$

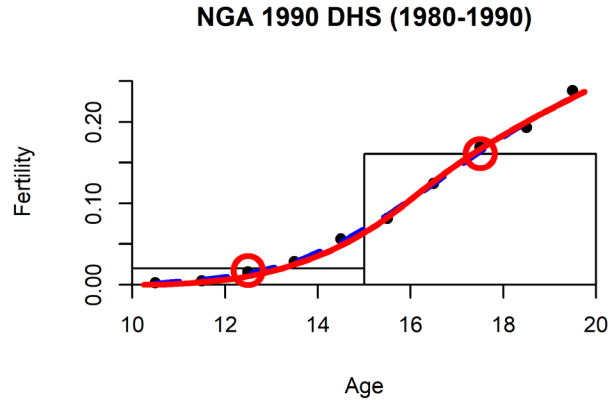


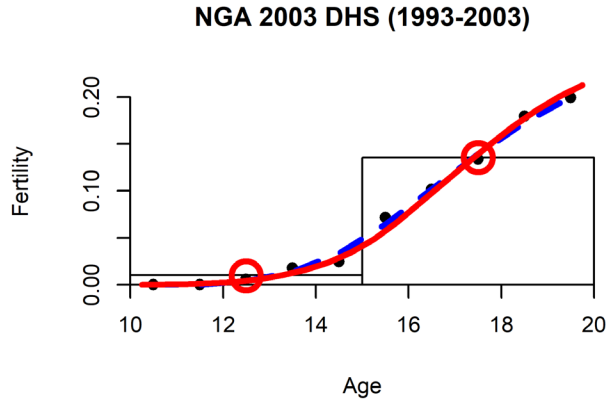
95% of countries have RMSE < 0.0022

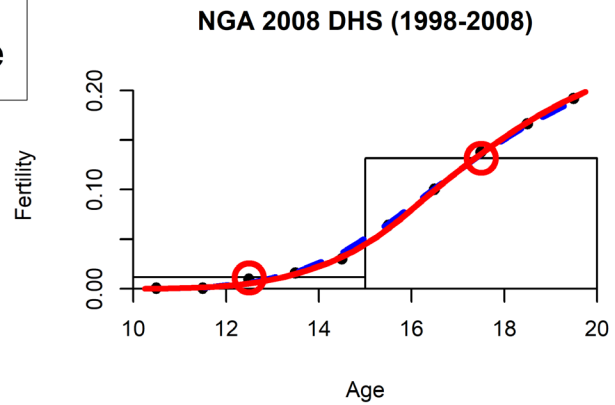
Australia and New Zealand
Central and Southern Asia
Eastern and South-Eastern Asia
Europe and Northern America
Latin America and the Caribbean
Northern Africa and Western Asia
Oceania*
Sub-Saharan Africa

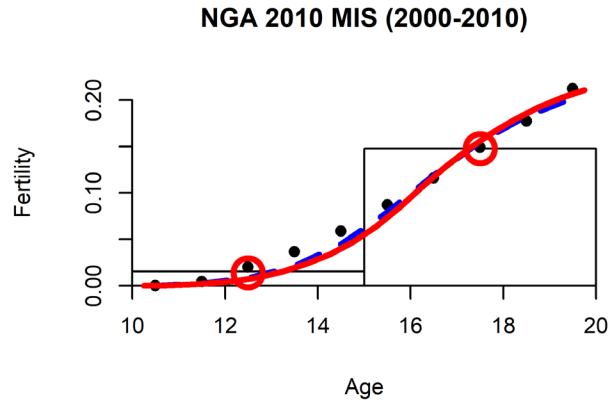

Observed
Smoothed single age
5-year age average
Estimated single age

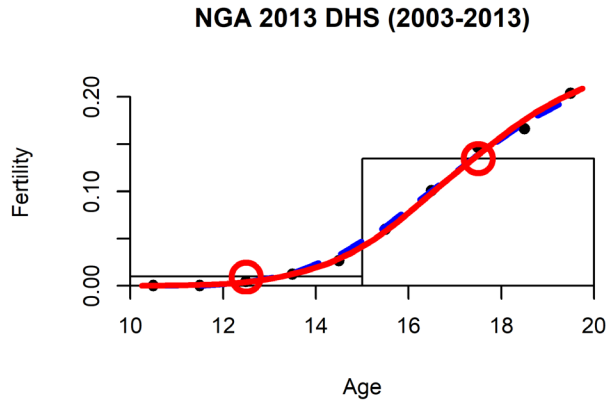

NGA 1981-1982 WFS (1972-1982)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age

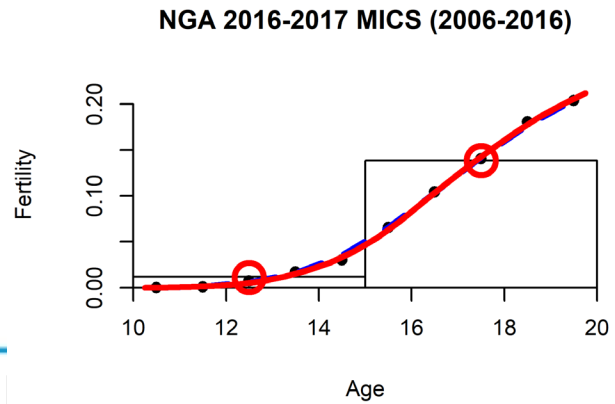

NGA 1990 DHS (1980-1990)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age

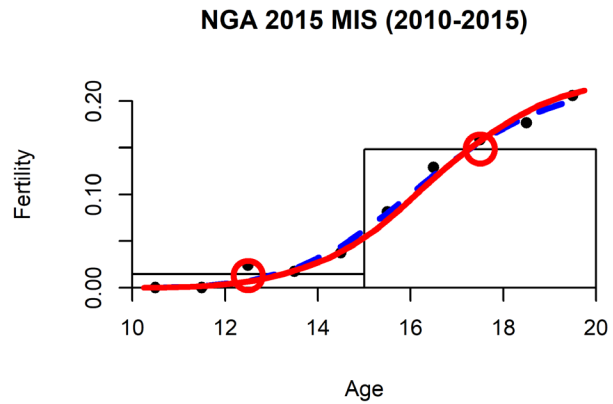

NGA 2003 DHS (1993-2003)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age

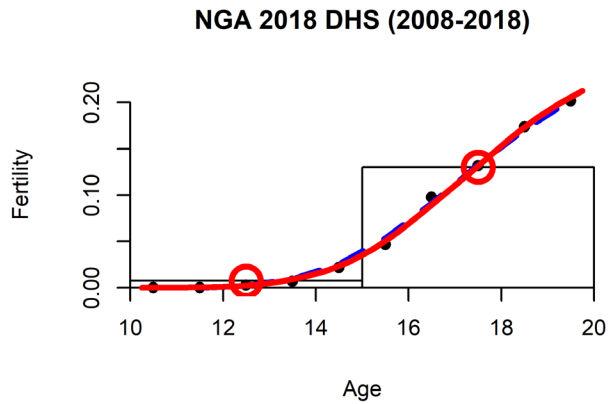

NGA 2008 DHS (1998-2008)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age


NGA 2010 MIS (2000-2010)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age

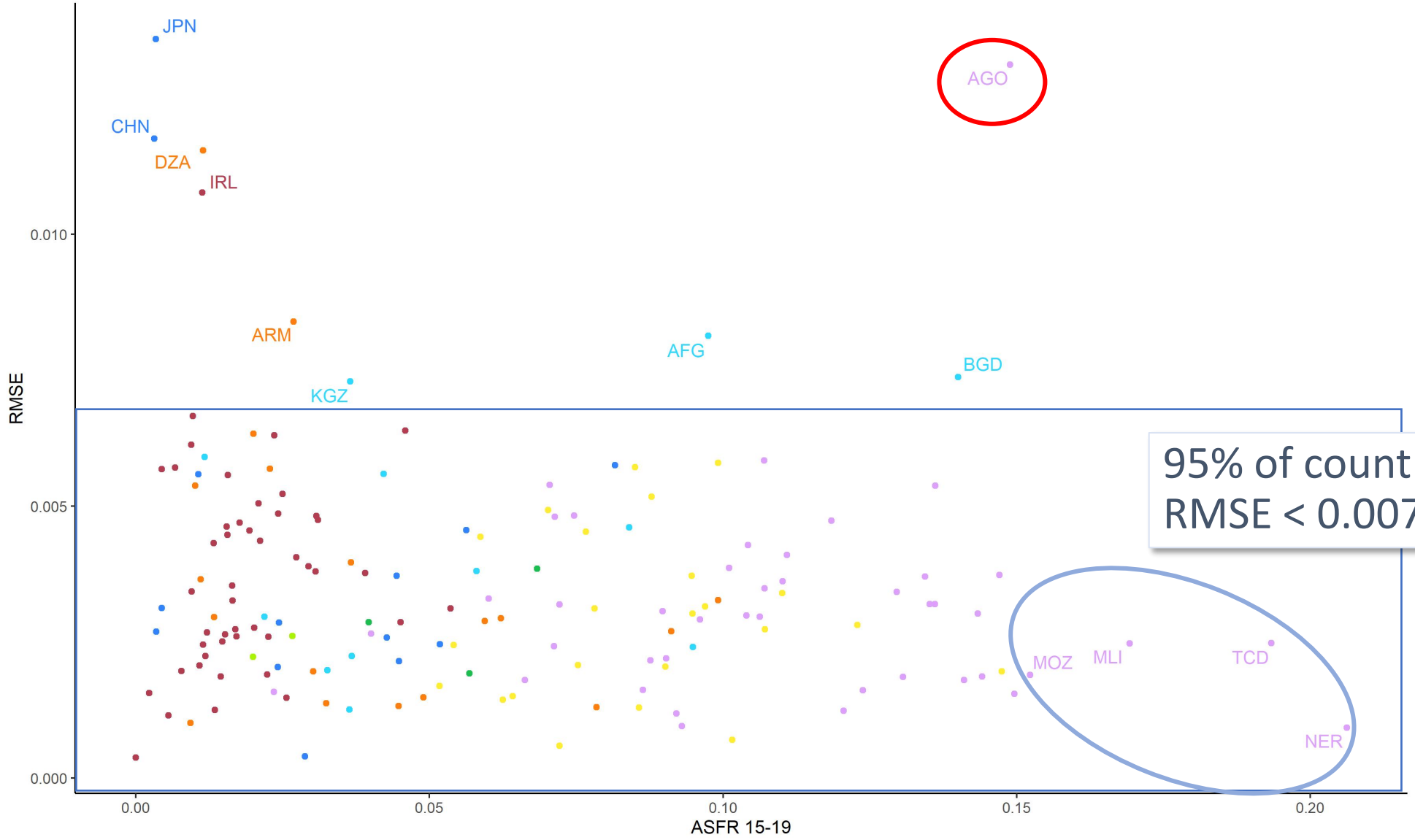

NGA 2013 DHS (2003-2013)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age


NGA 2016-2017 MICS (2006-2016)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age


NGA 2015 MIS (2010-2015)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age

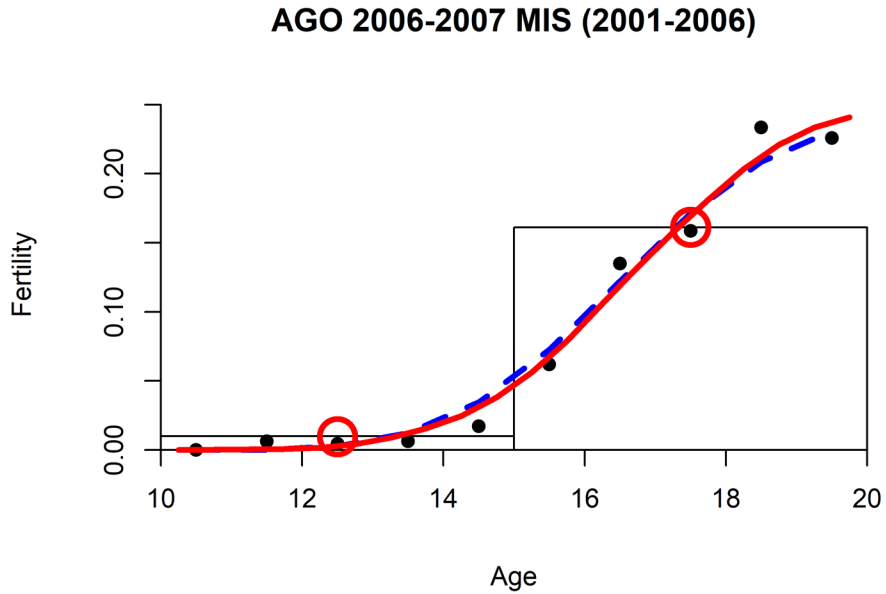

NGA 2018 DHS (2008-2018)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age

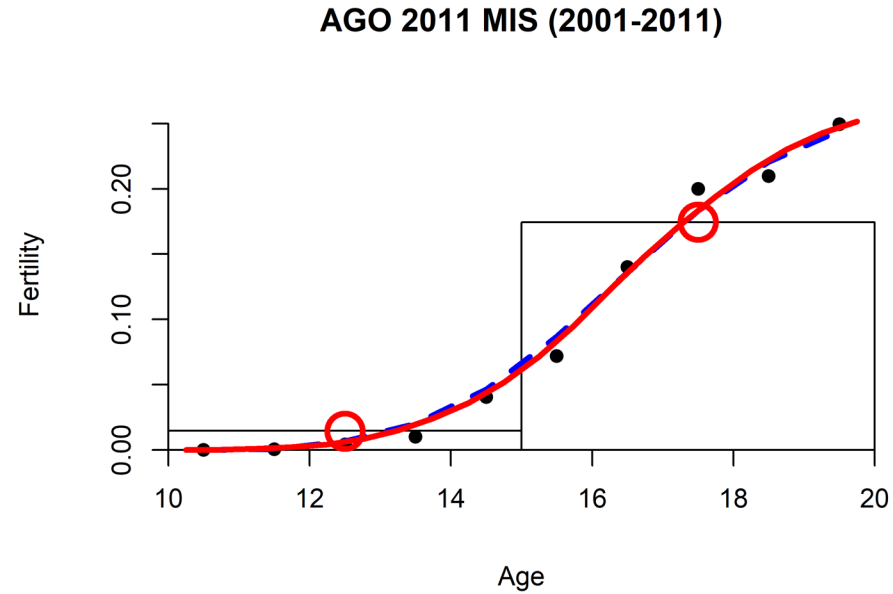
Model performance on age 15-19 by country

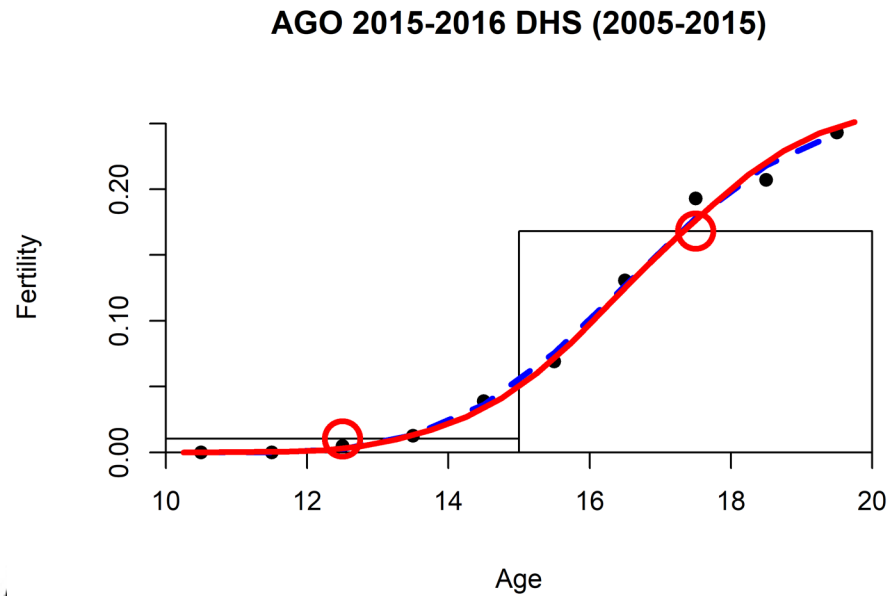


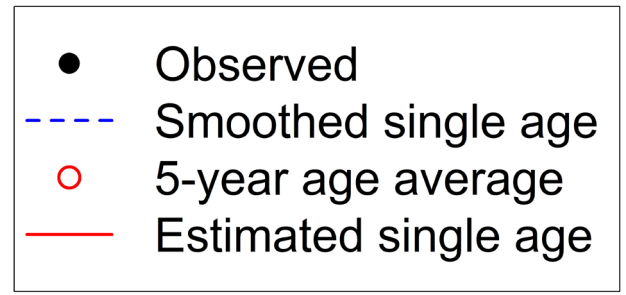
95% of countries have RMSE < 0.007

Australia and New Zealand
Central and Southern Asia
Eastern and South-Eastern Asia
Europe and Northern America
Latin America and the Caribbean
Northern Africa and Western Asia
Oceania*
Sub-Saharan Africa


AGO 2006-2007 MIS (2001-2006)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age


AGO 2011 MIS (2001-2011)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age


AGO 2015-2016 DHS (2005-2015)
Fertility
0.00
0.10
0.20
10
12
14
16
18
20
Age


Observed
Smoothed single age
5-year age average
Estimated single age

Conclusions for fertility age patterns

- More than 4,500 series with single age fertility available from VR, surveys and HDSS for 171 countries, covering age 10-54 from 1891 onwards,
- VR series used as-is for countries/areas with complete birth registration and accurate reporting of the age of mother (i.e. no heaping)
- For other countries/areas, use 5-year age groups graduated into single age using a recalibrated spline model developed by Schmertmann:
  - VR data + Survey/HDSS series smoothed using PCA and cubic spline
  - Fertility multipliers compiled using a calibrated spline estimator. The model:
    - is able to reproduce single age fertility patterns when only 5-year age are available,
    - works well across different regions and time, and with different shapes of fertility distribution by age,
    - Provides reasonable estimates also for fertility at very young ages (10-14) and old ages (50+)

References

- Pantazis A, Clark S.J. (2018). A parsimonious characterization of change in global age-specific and total fertility rates. *PLoS ONE 13 (1)*: e0190574. https://doi.org/10.1371/journal.pone.0190574
- Schoumaker, B. (2020). Presentation for the United Nations Expert Group Meeting on methods for the World Population Prospects 2021 and beyond. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/unpd_egm_202004_s2_schoumaker.pdf
- Schmertmann, C. P. (2014). Calibrated spline estimation of detailed fertility schedules from abridged data. *Revista Brasileira de Estudos de População, 31(2)*, 291-307. https://www.scielo.br/scielo.php?pid=S0102-30982014000200004&script=sci_arttext